

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE CIÊNCIAS BÁSICAS DA SAÚDE  
CURSO DE GRADUAÇÃO EM BIOMEDICINA

JULIANO LUIZ FACCIÓNI

**DESENVOLVIMENTO DE FERRAMENTA DE RASTREAMENTO  
AUTOMATIZADO DE CÉLULAS E EXPLORAÇÃO DE EVENTOS DE  
HETEROGENEIDADE EM ÁRVORES DE DESCENDÊNCIA**

Porto Alegre  
Dezembro de 2018

JULIANO LUIZ FACCIÓNI

**DESENVOLVIMENTO DE FERRAMENTA DE RASTREAMENTO  
AUTOMATIZADO DE CÉLULAS E EXPLORAÇÃO DE EVENTOS DE  
HETEROGENEIDADE EM ÁRVORES DE DESCENDÊNCIA**

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul, como requisito parcial para obtenção do título de Bacharel em Biomedicina.

Orientador: Prof. Dr. Guido Lenz

Porto Alegre  
Dezembro de 2018

## AGRADECIMENTOS

À Universidade Federal do Rio Grande do Sul, pelo apoio institucional e pela formação acadêmica e científica que recebi de seu brilhante corpo docente durante meus anos de graduação;

Ao professor e orientador Dr. Guido Lenz, pelo encorajamento contínuo, pela orientação única e por instigar em seus alunos o desafio fundamental de encarar a ciência sob um novo olhar;

Aos colegas do LabSinal, pelo companheirismo e troca científica realizada todos os dias entre bancadas e reuniões. Em especial à Giovana, que me orientou com paciência e disponibilidade durante meu período de iniciação científica e contribuiu imensamente para minha formação;

Aos amigos de infância, amigos de Porto Alegre, colegas de biomedicina e àqueles espalhados pelo Brasil ou até mesmo pelo mundo, por fazer parte de minha vida e celebrá-la em minha companhia;

À Laura, minha namorada, pela compreensão, companheirismo e apoio irrestrito nesta caminhada ao meu lado;

À minha família: minhas irmãs Alessandra e Vanessa e meus pais Jorge e Vânia, pela educação e apoio incondicional durante toda a trajetória de minha vida; e também ao meu sobrinho Augusto, pelos momentos de leveza e felicidade.

## SUMÁRIO

<b>1 RESUMO.....</b>	<b>4</b>
<b>2 INTRODUÇÃO.....</b>	<b>5</b>
<b>2.1 Células únicas.....</b>	<b>5</b>
2.1.1 O paradigma de célula única.....	5
2.1.2 Células únicas e populações celulares.....	5
2.1.3 Metodologias para células únicas.....	6
<b>2.2 De células únicas a colônias celulares.....</b>	<b>8</b>
2.2.1 Células irmãs e heterogeneidade genealógica.....	8
2.2.2 Colônias celulares e evolução clonal.....	9
<b>2.3 Desafios na análise de imagens de células únicas.....</b>	<b>12</b>
2.3.1 Viabilidade celular e resolução temporal.....	12
2.3.2 Volume e tratamento de dados.....	12
2.3.3 Necessidade de automatização.....	13
<b>2.4 Automatização como aliada a análises de células únicas.....</b>	<b>16</b>
2.4.1 Programas abertos à comunidade científica.....	16
2.4.2 Desafios na análise dos dados gerados.....	17
<b>3 OBJETIVOS.....</b>	<b>19</b>
<b>3.1 Justificativa.....</b>	<b>19</b>
<b>3.2 Objetivo Geral.....</b>	<b>19</b>
<b>3.3 Objetivos Específicos.....</b>	<b>19</b>
<b>4 REDAÇÃO CIENTÍFICA.....</b>	<b>20</b>
<b>FATE, a tool for automated cell tracking and lineage exploration.....</b>	<b>21</b>
<b>Methods.....</b>	<b>29</b>
Experimental setup.....	29
Image analysis.....	30
FATE analysis.....	30
<b>References.....</b>	<b>31</b>
<b>5 PERSPECTIVAS FUTURAS.....</b>	<b>33</b>
<b>6 REFERÊNCIAS DO TEXTO PRINCIPAL.....</b>	<b>35</b>
<b>7 ANEXOS.....</b>	<b>39</b>
<b>7.1 Normas de publicação para Nature Methods – Brief Communication.....</b>	<b>39</b>



# 1 RESUMO

A heterogeneidade inerente a virtualmente todas as populações celulares carrega importantes implicações para doenças como o câncer, onde uma pequena população fenotipicamente distinta pode ser indetectável por metodologias analíticas populacionais mas conter grande relevância terapêutica caso componha uma população resistente a quimioterápicos. Experimentos com células únicas são capazes de gerar dados mais robustos do que análises populacionais, permitindo a detecção de subgrupos raros e um detalhamento mais fino da heterogeneidade de uma população. Entretanto, estes experimentos geram um volume considerável de dados, os quais são inviáveis de ser quantificados de maneira manual. Muitos programas foram desenvolvidos em uma tentativa de automatizar a detecção de células únicas em experimentos contínuos de microscopia; contudo, a análise final dos dados gerados ainda é complexa e requer conhecimentos técnicos de métodos computacionais. Neste trabalho, apresentamos o programa Family Tree Explorer (FATE). O programa tem como principal objetivo facilitar a análise de dados posterior a uma detecção automatizada de células em sequências de imagens de microscopia. FATE é capaz de rastrear células únicas através da sequência de fotos e inferir a ocorrência de divisão celular, migração e morte. Além disso, FATE consegue construir em memória a história de cada células e de sua árvore de descendentes. No modelo do programa, cada célula em cada árvore criada carrega informações únicas sobre si, como por exemplo tamanho celular, intensidade de fluorescência e referência em memória para suas células filhas/irmã/mãe. O programa permite explorar relações entre estas informações e avaliar a heterogeneidade em populações celulares de maneira inovadora, em análises voltadas tanto para comparações entre as células de uma dada árvore quanto para árvores distintas entre si. A partir das análises realizadas pelo FATE, espera-se colaborar com o surgimento de novas metodologias de análise de árvores de descendentes, bem como ajudar os cientistas a automatizar e simplificar seus protocolos de análise de células únicas em imagens de microscopia.

**Palavras-chave:** células únicas; evolução clonal do câncer; automatização; análise de imagens.

## 2 INTRODUÇÃO

### 2.1 Células únicas

#### 2.1.1 O paradigma de célula única

Populações celulares são heterogêneas. Este fato já é inferido e até mesmo conhecido empiricamente por cientistas há muito tempo (Sirachý, 1979; Woodruff, 1983), sendo de especial relevância para tumores (Hirsch et al., 1983), os quais muitas vezes são identificados como massas celulares homogêneas a olho nu. Entretanto, a análise e quantificação do quão heterogênea uma dada população de células realmente é tornou-se possível apenas com os recentes avanços de tecnologias como microarranjos de RNA, microfluídica, marcação fluorescente de proteínas, bem como com a maior capacidade computacional de armazenamento e processamento de dados (Spiller et al., 2010). Estas análises são realizadas através da medida de parâmetros de cada uma das células de uma população; são, portanto, conjuntamente denominadas de *análises de células únicas*.

#### 2.1.2 Células únicas e populações celulares

A percepção de que células únicas diferem de sua população é, de certa forma, uma conclusão óbvia. Afinal, um dos conceitos básicos de estatística é justamente a ideia de que indivíduos de uma população não são idênticos à sua média populacional. A questão crucial passa a ser, portanto, não somente demonstrar que células aparentemente idênticas são diferentes, mas identificar *quando* elas são diferentes, *o quanto* diferem e *em que situações* esta diferença realmente carrega um significado biológico relevante além de um ruído estocástico (Altschuler & Wu, 2010). Ainda assim, é importante ressaltar algumas vantagens que as abordagens de células únicas apresentam sobre análises populacionais: a descrição mais precisa e mais detalhada da complexidade clonal de células em uma população (Wills & Mead, 2015); a identificação de subpopulações celulares pequenas mas de alta relevância biológica, como bactérias resistentes a antibióticos (Balaban et al., 2004) ou células tumorais capazes de sobreviver ao tratamento quimioterápico e repopular o tumor (Tellez-Gabriel et al., 2016); e o desenvolvimento de metodologias capazes de gerar novas descobertas e *insights* científicos, como o estudo detalhado de células irmãs (Luyi et al., 2018) ou de evolução tumoral (Navin et al., 2011), que simplesmente não são realizáveis ao se trabalhar com populações celulares.

### 2.1.3 Metodologias para células únicas

Na maioria dos casos, as técnicas aplicadas para análises de células únicas são similares às existentes para populações celulares. O grande diferencial está no avanço da técnica em si, permitindo análises em células únicas de características até então apenas observadas em populações. Obviamente, a quantidade de dados gerados com análises de células únicas é muito superior aos obtidos por análises populacionais.

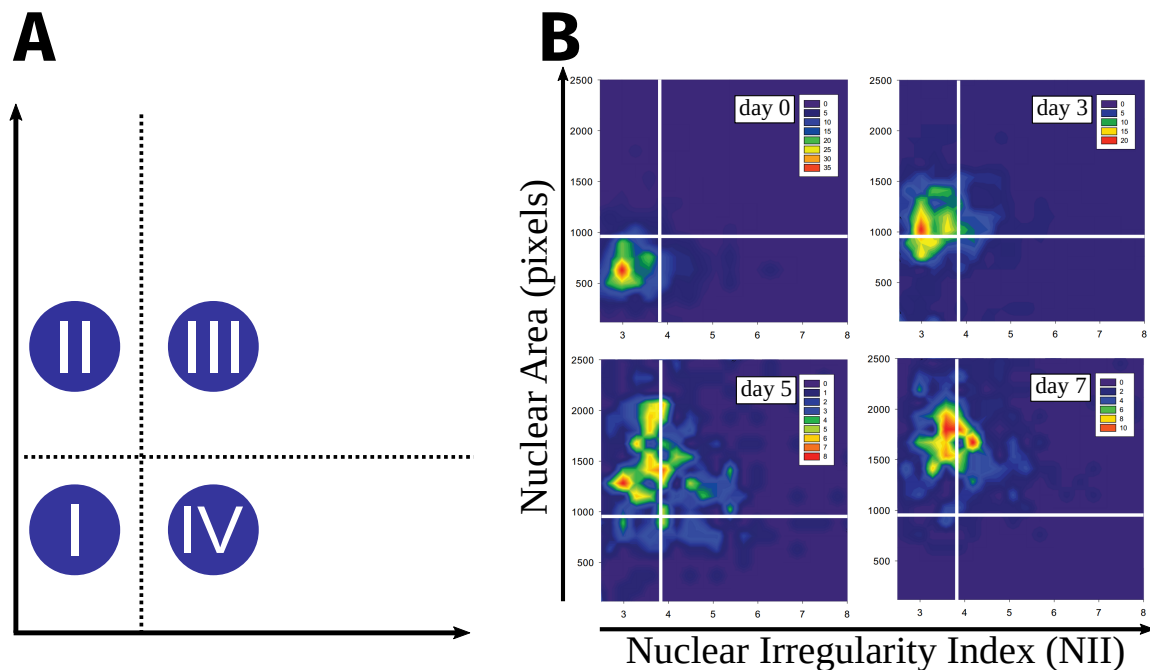
Algumas das principais tecnologias utilizadas para a análise de células únicas são as *ômicas*: genômica, transcriptômica e proteômica. Tratam-se de técnicas relativamente modernas, originárias de análises populacionais; contudo, o interesse de análises ômicas em células únicas tem crescido em um esforço que segue ao encontro da importância de se analisar células únicas em uma população (Wang & Bodovitz, 2010). O desafio de sequenciar o DNA, RNA ou identificar proteínas expressas em células individuais só foi transposto a partir do desenvolvimento de tecnologias mais sofisticadas de microfluídica e análise computacional dos dados gerados, como, por exemplo, reações de PCR ocorrendo dentro de gotículas em capilares microfluídicos, ou o *barcoding* de células únicas, permitindo estabelecer-se *in silico* a relação entre material genético amplificado e célula de origem (Nawy, 2013).

Outra grande aliada na análise de células únicas é a microscopia de fluorescência. Apesar de ser uma técnica já bem estabelecida, o grande avanço no sentido de analisar células individuais com algum marcador fluorescente (seja via imunohistoquímica, transfecção/transdução vetorial, ou ainda por tecnologias mais recentes como CRISPR) vem, principalmente, da computação. A capacidade de analisar grandes volumes de imagens de fluorescência de maneira automatizada e padronizada permite que os dados resultantes passem de uma análise meramente qualitativa para quantitativa e, em muitos casos, multidimensional (Mattiuzzi Usaj et al., 2016). É possível quantificar, de maneira mais eficiente e objetiva, os componentes e processos celulares relacionados aos marcadores fluorescentes. Por exemplo, o plasmídeo FUCCI é capaz de reportar o estágio do ciclo celular de cada célula através de marcadores fluorescentes fusionados a proteínas diferencialmente expressas em determinadas etapas do ciclo celular (Sakaue-Sawano et al., 2008); já o plasmídeo mCherry-GFP-LC3B, cujo produto de expressão é recrutado para a membrana de autofagossomos, serve como indicador da etapa do processo autofágico em função da

fluorescência relativa de seus marcadores (Hundeshagen et al., 2011). De fato, trata-se de uma análise tão comum e efetiva que diversos softwares e metodologias de alto rendimento foram criados para facilitar e agilizar o processamento de imagens de fluorescência (Eliceiri et al., 2012).

Uma tecnologia mais recente para análise de células únicas é a citometria de massa, que combina a detecção de alta resolução da espectrometria de massa com a capacidade da citometria em analisar uma célula por vez. Utilizando a relação entre massa de anticorpos diferencialmente ligados a metais de transição, é possível obter uma quantificação precisa e multidimensional de mais de 20 marcadores para cada célula individual de uma população (Ornatsky et al., 2010).

É importante reforçar que análises de células únicas também são capazes de descrever a população a partir de cada célula que a compõe. O NMA (análise morfométrica nuclear, do inglês *nuclear morphometrical analysis*) usa a descrição de características morfológicas do núcleo de células para classificá-las em diferentes processos celulares (células regulares, senescentes, apoptóticas e com irregularidades nucleares) e, assim, gerar uma distribuição populacional de estados celulares (Filippi-Chiela et al., 2012) (figura 1).



**Figura 1. Apresentação da técnica de NMA.** **A**, representação do gráfico gerado por NMA. As regiões destacadas representam distintas populações celulares: I, células regulares; II, células senescentes; III, células apoptóticas; IV, células com irregularidades nucleares (potencialmente passando por catástrofe mitótica). **B**, *Heatmaps* representativos de resultados de NMA. Células da linhagem de glioblastoma U87 foram tratadas com 50  $\mu\text{M}$  de temozolomida por 3 h. Marcações nucleares com DAPI foram realizadas nos intervalos indicados, imagens dos núcleos marcados foram adquiridas por microscopia de fluorescência e as imagens foram analisadas pelo *plugin* de NMA para o software aberto ImageJ. 2015, dados não publicados de Oliveira, A; Lenz, G.

## 2.2 De células únicas a colônias celulares

### 2.2.1 Células irmãs e heterogeneidade genealógica

Em muitos experimentos com células únicas, o interesse não está apenas em realizar uma medida pontual de todas as células de uma população, mas sim diversas medidas temporalmente espaçadas de cada célula única. Isto é o que ocorre, por exemplo, em fotografias de *time-lapse* de alto rendimento. Nestas situações, a grande maioria das células únicas eventualmente passará pelo processo de divisão celular e gerará duas células filhas.

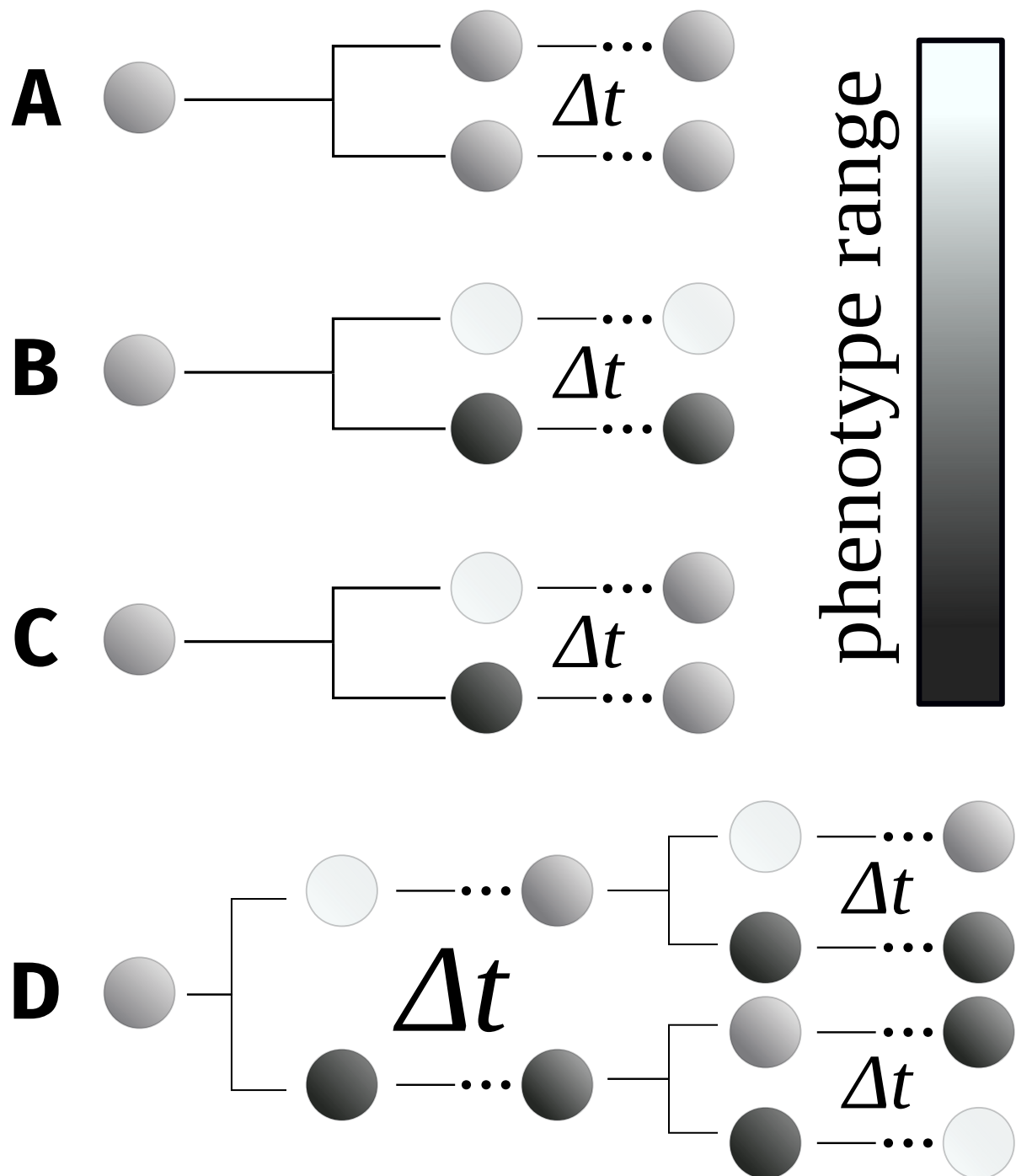
É possível imaginar que um processo tão complexo e delicado como a mitose dificilmente resultará em duas células filhas idênticas; de fato, há diversos exemplos na biologia que demonstram a assimetria da divisão celular (Lechler and Fuchs, 2005; Strome and Wood, 1983; Yamashita et al., 2003). Por exemplo, existem evidências de que células que apresentam aspectos de células-tronco (*stemness*) são capazes de segregar suas organelas de maneira

assimétrica, passando as organelas mais antigas (e portanto mais propensas a apresentar acúmulo de dano) para uma de suas filhas e preservando o estado de *stemness* da outra filha (Katajisto et al., 2015). Isto é relevante ao se considerar o ponto de vista de uma célula-tronco, a qual deve, simultaneamente, gerar um descendente capaz de se diferenciar em algum tipo celular mais especializado, e outro capaz de continuar com a capacidade de autorrenovação característica de uma célula-tronco.

O principal conceito que guia análise de células únicas – a ideia de que células únicas diferem da média de sua população e guardam informações relevantes, porém imperceptíveis ao se analisar a população como um todo – pode ser expandido para as suas descendentes, para as descendentes dessas descendentes, e assim por diante. Esta perspectiva pode ser descrita pelo termo *heterogeneidade genealógica* (figura 2B); isto é, células únicas geram descendentes únicos, diferentes tanto em comparação com a célula irmã originada da mesma célula mãe, quanto em comparação com os demais descendentes da população. Este conceito, ainda que não necessariamente inovador (Chatterjee and Acar, 2018; Kester and Oudenaarden, 2018), é pouco explorado. Por exemplo, sabe-se que alterações no processo de divisão assimétrica podem ser responsáveis pelo surgimento de tumores (Morrison and Kimble, 2006), porém se desconhece a relação (ou a ausência dela) entre a heterogeneidade genealógica e a progressão tumoral e/ou resistência de um tumor à terapia. Novos estudos nesta área podem, potencialmente, gerar impactos científicos semelhantes àqueles provocados pela mudança do paradigma de análises populacionais para análises de células únicas.

### **2.2.2 Colônias celulares e evolução clonal**

Dado o tempo necessário, uma célula única se replicará gerando duas filhas. Cada filha eventualmente gerará duas novas filhas, e assim por diante, culminando no surgimento de uma colônia celular. Esta população apresenta grande semelhança entre seus indivíduos justamente por ter se originado de uma célula única; é, portanto, uma *população clonal* (figura 2A).



**Figura 2. População clonal, heterogeneidade genealógica e evolução clonal.** **A**, em uma população puramente clonal, as células filhas são cópias idênticas de sua mãe e não apresentam variações fenotípicas como função do tempo. **B**, a heterogeneidade genealógica é responsável pelas diferenças fenotípicas observadas para células filhas de uma mesma mãe. **C**, o fenótipo das células filhas pode variar não somente em decorrência da heterogeneidade genealógica, mas também em função do tempo. **D**, a evolução clonal é a descrição mais detalhada da maneira com que células únicas variam seu fenótipo dentro de uma população. A variabilidade é introduzida na população tanto pelo fator mitótico (heterogeneidade genealógica) quanto pelo fator temporal, sendo propagada através de gerações de células únicas.

O conceito de população clonal é de extrema relevância para a biologia do câncer. O consenso moderno afirma que grande parte dos tumores se origina de populações de poucas células – possivelmente de uma célula única – as quais sofrem mutações aleatórias e passam a apresentar valor adaptativo (*fitness*) maior que as demais em sua população (Ding et al., 2013; McGranahan and Swanton, 2017). Esta célula, bem como suas descendentes, terão mais sucesso em competir por recursos, replicar-se, evadir células do sistema imune e frear processos celulares desfavoráveis como a apoptose. Por outro lado, acredita-se que as células responsáveis por recolonizar um tumor podem ser originadas de um subclone já existente no tumor antes do tratamento. Este subclone resiste ao tratamento (seja por mecanismos genéticos ou epigenéticos) e é capaz de recolonizar a região, causando a recidiva do câncer (Landau et al., 2014).

Surge aqui uma aparente contradição entre os conceitos de população clonal e heterogeneidade genealógica: células únicas tendem a gerar filhas mais semelhantes a si do que às demais células da população; por exemplo, há indicativos de que diferenças no desfecho celular podem ser herdáveis mesmo em células pertencentes a uma mesma população clonal (Spencer et al., 2009). Contudo, estas células filhas podem se distanciar genética e fenotipicamente de sua mãe, como consequência da heterogeneidade genealógica. Uma questão relevante, portanto, é: *o quão* semelhantes células filhas são em relação à sua mãe, e *por quanto tempo* dura esta semelhança? O conceito que tenta relacionar estes processos é denominado de *evolução clonal* (figura 2D).

A evolução clonal tem ganhado destaque em estudos sobre tumores, em especial a partir do surgimento de metodologias computacionais capazes de classificar células únicas em *clusters* de evolução clonal, com o intuito de desenvolver uma medicina personalizada para os subtipos tumorais de cada paciente. Embora ainda distantes da realidade sob um ponto de vista prático, os esforços neste sentido são diversos e contínuos (Ding et al., 2013). O principal desafio para os cientistas da área biomédica deixa de ser as possíveis limitações técnicas ou o esforço manual na execução de protocolos, e passa a ser a interpretação correta e imparcial do enorme volume de dados gerados e a extração de conclusões válidas e relevantes para a área.



## **2.3 Desafios na análise de imagens de células únicas**

### **2.3.1 Viabilidade celular e resolução temporal**

A principal limitação em relação à captura de imagens de células únicas durante prolongados períodos é garantir que as características fisiológicas das células se mantenham inalteradas pelas próprias condições experimentais. É preciso manter os níveis de CO<sub>2</sub>, umidade e temperatura adequados, evitar a fototoxicidade e garantir uma boa acessibilidade óptica para que imagens possam ser capturadas. Além disso, dependendo da natureza do experimento, pode ser necessário o acesso físico ao meio de cultivo, ou uma superfície de cultivo específica.

Apesar da grande gama de soluções comerciais disponíveis, não existe nenhum equipamento capaz de adequar-se a todas estas requisições (Skylaki et al., 2016), e os que melhor se aproximam delas requerem um investimento financeiro significativo e em muitos casos indisponível. Uma alternativa encontrada por muitos laboratórios é uma “semiautomatização” da captura de fotos; isto é, manter as células em cultivo na estufa e, em intervalos regulares de tempo, utilizar um equipamento de captura automatizada de imagens pontuais. Até certo ponto, isto contorna alguns problemas como a manutenção dos níveis de CO<sub>2</sub> e temperatura, uma vez que as células permanecem por pouco tempo fora da estufa. A desvantagem evidente deste método é a resolução temporal limitada do experimento, além de requerer um esforço manual para levar as células da estufa para o equipamento e vice-versa. Nestas circunstâncias, um desenho experimental envolvendo imagens a cada 30 minutos durante duas semanas torna-se simplesmente inviável.

### **2.3.2 Volume e tratamento de dados**

O volume de dados gerados por experimentos imagéticos de células únicas não deve ser subestimado. Considerando o seguinte desenho experimental demonstrativo: células plaqueadas em placa de 96 poços, captura de 9 campos por poço, imagens realizadas a cada uma hora, duração do experimento de duas semanas. Ao final do experimento, são geradas 290.304 imagens. Supondo que cada imagem tenha uma resolução adequada para uma análise delicada de intensidade (ocupando cerca de 6 megabytes de espaço em disco), o experimento gera um volume de dados de, aproximadamente, 1,7 terabytes.

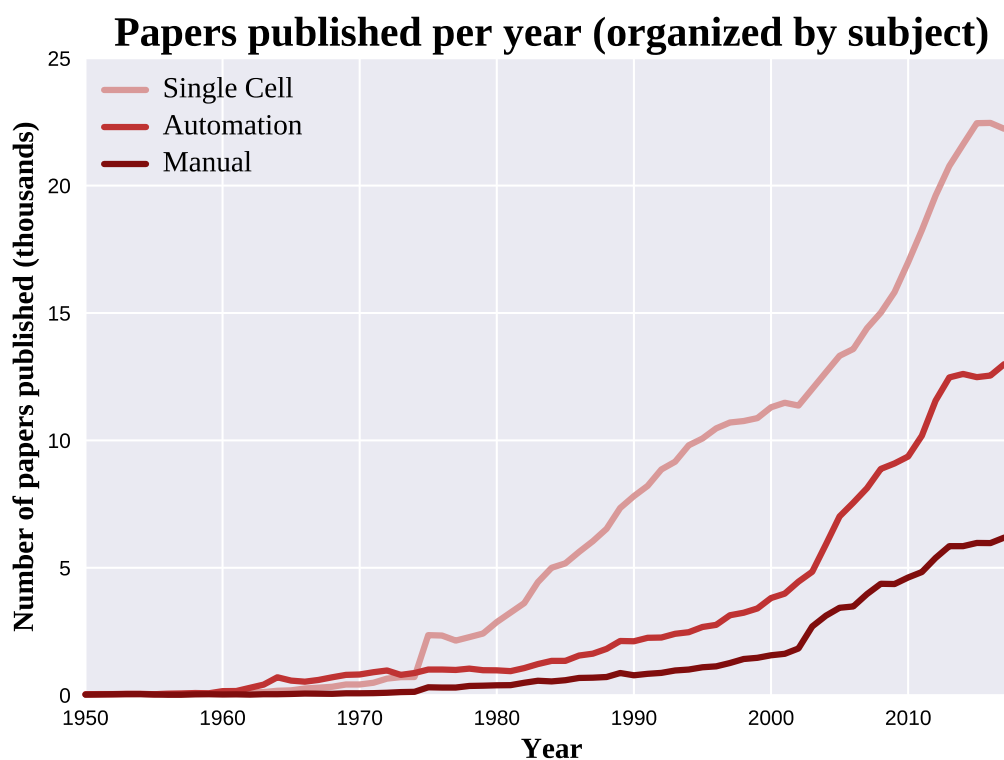
Apesar de ser uma premissa básica para a disponibilidade e compartilhamento de dados, pouca atenção é dada para a maneira como se armazena dados experimentais. Este fato contribui para a “entropia de dados”: com o passar do tempo, dados se tornam menos disponíveis para consulta, devido a deleções, corrompimento e formatações (Hart et al., 2016). Embora existam soluções comerciais da ordem de petabytes para armazenamento de dados a longo prazo, a maioria dos laboratórios não dispõe de recursos financeiros e tampouco de uma plataforma clara e adequada para o armazenamento de dados em larga escala (Skylaki et al., 2016). Sendo assim, o armazenamento de dados gerados a partir de imagens de células únicas compõe uma barreira para grupos de pesquisa que não contam com, por exemplo, servidores remotos para armazenamento.

Mesmo ignorando a questão do volume de dados, nem sempre é possível fazer a translocação dos dados gerados pela microscopia a análises informativas. Em muitos casos, o software responsável pela aquisição de imagens não é o mesmo que realiza a análise posterior (e por vezes não se trata nem do mesmo cientista). Algumas medidas simples que podem facilitar drasticamente a qualidade da análise – correção de iluminação de fundo, tempo de exposição adequado, escolha de nome de arquivo adequado para posterior extração de metadados – são decididas durante a aquisição. Cabe ao cientista manipulando o microscópio conhecer quais parâmetros devem ser ajustados para que a análise seja efetiva. Usuários avançados também são capazes de criar soluções próprias para ajustar a interface entre aquisição e análise, ou integrar plataformas de código aberto em seu fluxo de trabalho para sobrepujar limitações; contudo isto requer conhecimentos de computação e programação nem sempre óbvios ou acessíveis.

### **2.3.3 Necessidade de automatização**

Com base nas questões discutidas acima, fica claro que a automatização é um ponto crucial na aquisição e análise de imagens de células únicas. Simplesmente não é viável analisar manualmente milhares de imagens, cada uma contendo dezenas ou centenas de células, tanto do ponto de vista de aproveitamento de tempo (dependendo do volume de dados, um humano poderia levar meses para analisar imagens uma a uma) quanto de produtividade: gastar tempo realizando tarefas repetitivas que poderiam ser automatizadas significa *não* gastar tempo com outras tarefas possivelmente mais produtivas, como formulação de novas hipóteses, execução de outros experimentos e escrita científica.

Isto vai ao encontro da tendência mundial de *automatização de processos manuais*, um fato cada vez mais presente não somente nas ciências biomédicas (Hawker, 2007; Novak and Marlowe, 2013; Riben, 2015; Konczal and Gray, 2017) ou no meio acadêmico em geral (Zavadskas, 2010; Banks et al., 2014; Correa-Baena et al., 2018) (figura 3), mas em virtualmente todas as esferas produtivas da sociedade moderna (Rotman, 2017). A automatização do próprio processo científico de realização de experimentos e formulação/teste de hipóteses já é uma realidade em alguns centros de pesquisa de ponta (King et al., 2009).

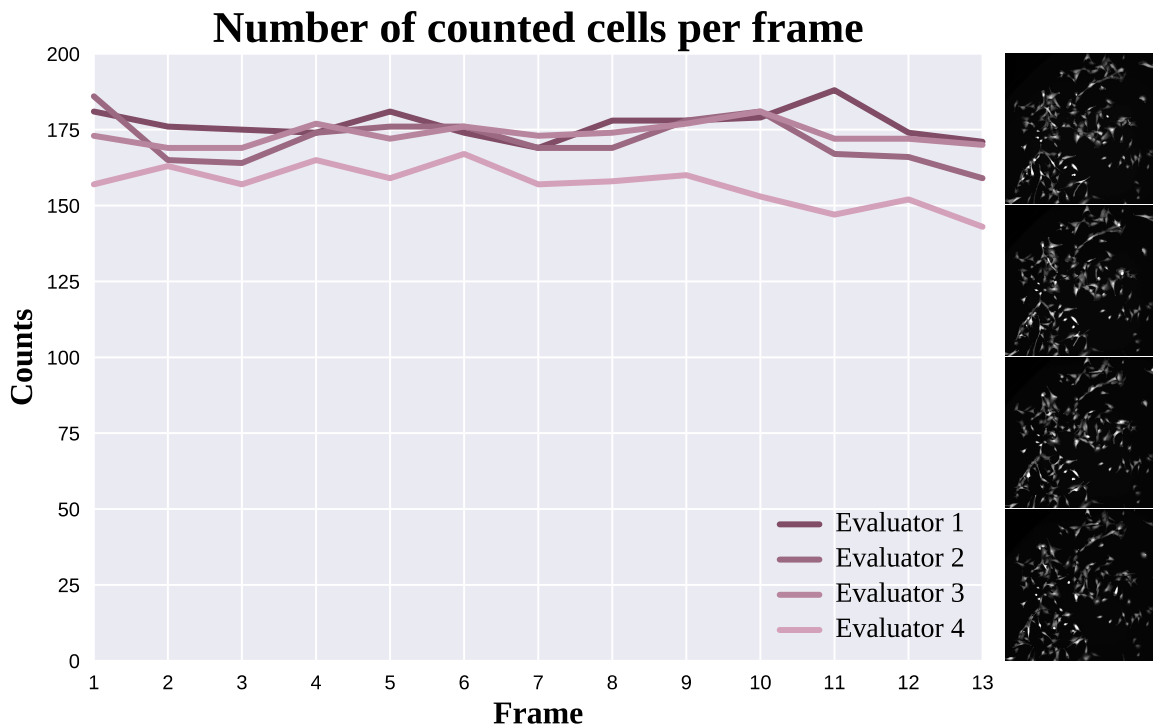


**Figura 3. Automação como tendência global.** Artigos publicados por ano (unidade em milhares) para cada tema indicado pela legenda. Os resultados foram obtidos através de uma busca simples na base de dados PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>), utilizando os temas indicados como palavra-chave. Dados obtidos em: 26 set 2018.

Um dos principais contrapontos empíricos à automatização é o investimento a médio prazo necessário para dominá-la. Qualquer tentativa de automatizar um processo, seja através da adoção de uma ferramenta já existente, ou pela criação de uma metodologia personalizada para atender a uma demanda própria, requer a transposição de uma barreira inicial de aprendizado. É preciso dedicar um tempo significativo para *aprender a automatizar* o

processo de interesse. Muitos cientistas preferem dedicar esforços em análises manuais capazes de trazer algum resultado a curto prazo, ainda que não satisfatório. É preciso conscientizar-se de que aprender a automatizar no presente poupará tempo e esforço no futuro. Há diversas ferramentas que tentam reduzir a barreira de entrada, trazendo uma interface mais intuitiva e ferramentas simplificadas; por outro lado, estes programas tendem a ser menos personalizáveis e generalizáveis para diferentes demandas.

Mesmo em situações onde o volume de dados é pequeno o suficiente para que análises manuais de células únicas como contagens, mensuração de área e intensidade, e marcação de regiões de interesse sejam factíveis, detecta-se o surgimento de flutuações inerentes à estocasticidade de qualquer processo manual (figura 4). Por outro lado, um programa de computador reage da mesma maneira perante um mesmo volume de dados; espera-se que um programa bem escrito gere o mesmo *output* sempre que lhe for entregue o mesmo *input* (fica implícita a necessidade de um controle de qualidade que garanta que o *output* gerado pelo programa corresponde à realidade). Desta maneira, a automatização de procedimentos é capaz de gerar não somente um ganho no tempo e na facilidade de análise, mas também na confiabilidade, reprodutibilidade e objetividade dos resultados obtidos, chegando em alguns casos a níveis de precisão comparáveis a peritos de uma dada área (Esteve et al., 2017). Em princípio, é sempre possível olhar para o programa e entender os passos lógicos tomados para chegar aos resultados encontrados (dado que o código-fonte esteja disponível), o que não pode ser dito para análises manuais de dados.



**Figura 4. Estocasticidade inerente a processos manuais repetitivos.** Comparação da contagem manual de células em 13 imagens sequenciais (as quatro primeiras representadas à direita). Cada linha representa a contagem realizada por um diferente avaliador. 2017, dados não publicados.

## 2.4 Automatização como aliada a análises de células únicas

### 2.4.1 Programas abertos à comunidade científica

Uma verdadeira coletânea de programas para análise de imagens têm surgido para auxiliar cientistas ao trabalhar com imagens de microscopia, desde soluções para aquisição e armazenamento de imagens, até automatização de diversas análises possíveis e visualização dos resultados obtidos (Eliceiri et al., 2012). As ferramentas podem ser pagas ou gratuitas, de código aberto ou não, generalistas ou específicas para casos restritos.

Um excelente exemplo de sucesso é o ImageJ, software concebido em 1987 e pioneiro na análise computadorizada de imagens, o qual hoje segue as diretrizes de um software de código aberto mantido pela comunidade científica (Schneider et al., 2012). Com seu foco em interoperabilidade e desenvolvimento baseado em necessidades reais de cientistas, o ImageJ apresenta diferentes versões, atualizações, *plugins* e módulos criados para atender a variadas demandas. A distribuição Fiji, por exemplo, foca em *plugins* para análises científicas na área

de ciências biomédicas (Schindelin et al., 2012); já a distribuição AstroImageJ é voltada para a análise de imagens de astronomia (Collins et al., 2017).

Outras soluções partiram do caso de sucesso do ImageJ para diversificar e facilitar as análises de imagens. O Icy, por exemplo, surgiu em 2011 como uma alternativa que incorpora o ImageJ e permite a manipulação de imagens através de sua interface intuitiva (Chaumont et al., 2012). O programa simplifica a organização e processamento de *stacks* de imagens, transformações como correção de cor e intensidade, e a criação de protocolos automatizados para tratamento de imagens em larga escala através de um editor de protocolo acessível para aqueles não fluentes em linguagens de programação.

Por fim, o CellProfiler, mantido desde 2006, permite ao usuário a análise em larga escala de imagens de células. Através da criação de um *pipeline* personalizado de processamento e análise de imagens, os usuários são capazes de analisar um grande gama de imagens de maneira idêntica e automatizada, gerando resultados que vão desde medidas celulares simples (contagem de células, tamanho celular) até ensaios morfológicos mais complexos (formato do núcleos, quantificação de organelas em relação à área celular, padrões proteicos em células únicas) (Carpenter et al., 2006). Trata-se, portanto, de uma estratégia viável para a análise automatizada de células únicas.

Vale constatar que a gama de programas disponíveis, embora intimidante em um primeiro momento, é fundamentalmente um ponto positivo para o desenvolvimento de uma ciência mais colaborativa e aberta. A existência de diferentes plataformas e ferramentas possibilita encontrar a solução mais adequada ao fluxo de trabalho de cada laboratório. Além disso, considerando a natureza de código aberto de muitos destes programas, usuários avançados são capazes de modificá-los para atender a uma demanda específica e potencialmente redistribuir o fruto de seu trabalho como um *plugin* ou recurso adicional da plataforma. De fato, o esforço conjunto das diferentes comunidades destes programas culminou com o lançamento, em 2018, do fórum online Image.sc ([www.image.sc](http://www.image.sc)), onde usuários de diferentes programas porém com objetivos em comum são capazes de trocar informações.

#### **2.4.2 Desafios na análise dos dados gerados**

O passo seguinte à identificação automatizada de células através das soluções descritas acima é a interpretação dos resultados obtidos e estruturação e tratamento dos dados gerados.

É necessário encontrar a maneira ideal para analisar os diversos parâmetros envolvidos, como intensidade de fluorescência representando diferentes compartimentos ou estruturas celulares, ou medidas morfológicas geradas pelos programas.

Se pensarmos na evolução clonal a que as células estão submetidas durante o experimento, esta questão torna-se ainda mais crucial. É necessário que haja uma forma de transformar simples descrições de parâmetros de objetos encontrados em imagens em uma representação fiel da complexidade e liberdade que células individuais apresentam ao se dividir, migrar, entrar em morte celular e mudar sua morfologia e seu fenótipo.

Não existe ainda uma ferramenta computacional que reúna esta funcionalidade de maneira simplificada. Em diversos casos, os cientistas são instruídos a realizar seu próprio tratamento de dados após o uso dos programas de identificação de células únicas em imagens. Isto se traduz na necessidade de aprender alguma linguagem de programação ou de trabalhar por longos períodos de tempo com tabelas, planilhas e bases de dados; no pior dos casos, a dificuldade e falta de compreensão quanto à maneira correta de tratar os dados leva a análises equivocadas e resultados incorretos.

Neste trabalho, criou-se uma ferramenta que facilita a etapa final de interpretação de dados em um protocolo de análise de imagens de células únicas. O foco da ferramenta é possibilitar uma forma simples de visualizar e explorar os dados gerados de uma maneira que reflita a real complexidade contida neste tipo de análise.

## 3 OBJETIVOS

### 3.1 Justificativa

Para muitos cientistas, a contagem, a identificação e o rastreamento de células únicas e de colônias celulares ainda é uma atividade manual, demorada e passível de erros e vieses. Há demanda por ferramentas capazes de automatizar este processo, analisando imagens de maneira precisa e rápida, e gerando informações compatíveis com a complexidade inerente a análises de células únicas. Esta ferramenta deve se basear em soluções de natureza aberta e comunitária já existentes e utilizá-las como alavanca inicial, a fim de evitar o desperdício de tempo ou esforços redobrados ao recriar algo já disponível.

### 3.2 Objetivo Geral

Desenvolver uma ferramenta computacional capaz de gerar árvores de parentesco a partir da identificação de células únicas em imagens temporalmente espaçadas, de modo a descrever objetivamente as histórias individuais de migração, mitose e desfecho celular de cada célula que compõe a população e de suas descendentes, e gerar informações de maneira automatizada, rápida e de larga escala.

### 3.3 Objetivos Específicos

- Criar uma ferramenta capaz de utilizar as identificações de células feitas pelo programa de código aberto CellProfiler como *input* para sua análise, integrando-se de maneira eficiente a um fluxo de trabalho automatizado;
- Testar o seu uso para análise de parâmetros de células únicas relevantes para o NMA, como por exemplo a ocorrência e frequência de divisão e morte celular, o tamanho nuclear e o desfecho celular;
- Explorar a capacidade da ferramenta em construir árvores de descendência de células únicas com potencial para analisar eventos de heterogeneidade genealógica e evolução clonal.



## 4 REDAÇÃO CIENTÍFICA

Manuscrito no formato requerido pelo periódico Nature Methods para conteúdo do tipo “Brief Communication” ([www.nature.com/nmeth](http://www.nature.com/nmeth)).

Autores:

- Juliano Luiz Faccioni ([julianofaccioni@gmail.com](mailto:julianofaccioni@gmail.com))
- Guido Lenz ([gulenz@gmail.com](mailto:gulenz@gmail.com))

Endereço correspondente:

- Laboratório de Sinalização e Plasticidade Celular, Departamento de Biofísica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil ([www.ufrgs.br/labsinal](http://www.ufrgs.br/labsinal)).

# **FATE, a tool for automated cell tracking and lineage exploration**

Faccioni JL<sup>1</sup>, Lenz G<sup>\*1,2</sup>

1 – Laboratory of Signalling and Cellular Plasticity, Department of Biophysics, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil

2 – Centre of Biotechnology, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil

**Data analysis is one of the most critical aspects for automating single cell imaging workflow. Here we describe FATE, a tool to automate data analysis of continuous single cell imaging using the output of the popular open source software CellProfiler. FATE is able to infer cell lineage, mitosis, migration and cell death, and reconstruct and explore the ancestral history of each single cell in each ancestral tree.**

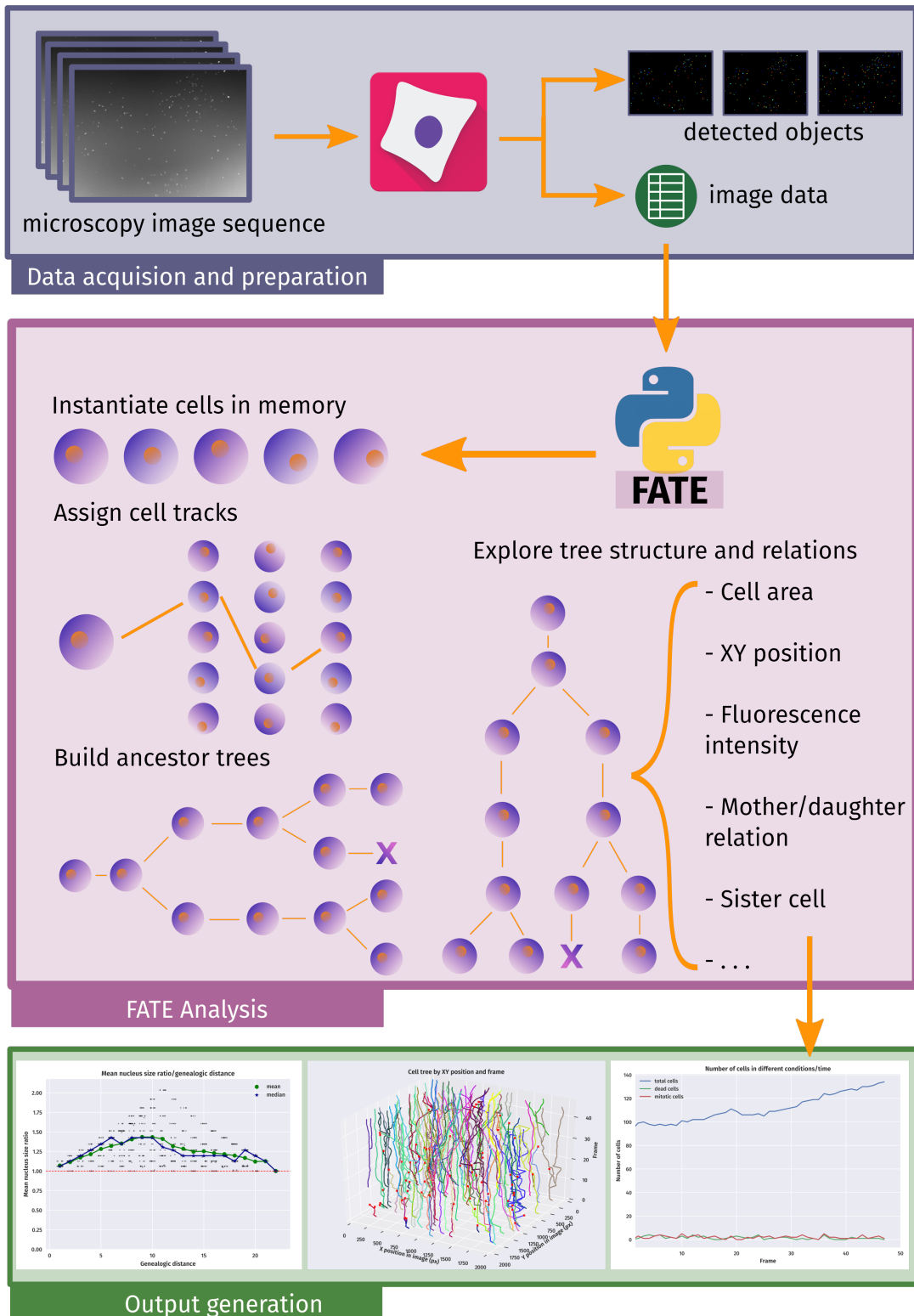
It is widely accepted that tumour heterogeneity – the fact that cancer cells from a same population may greatly differ genetically and phenotypically from their peers, both due to intrinsic genetic instability and to the complex signalling present in the tumour microenvironment – is one of the main factors responsible for the lethality of cancer and the emergence of drug-resisting clones in the tumour population<sup>1-4</sup>. The clonal evolution of cancer is a central theme in tumour therapy, and it is crucial to analyse it as an evolutionary process. Each tumour cell has an unique clonal evolutionary history, Darwinian fitness and drug response, with significant implications for cancer treatment<sup>5,6</sup>.

Single cell studies allow scientists to analyse cell populations as a collective of cell-to-cell measurements and variations, instead of a single mean value that does not take into account the nuances of each cell composing the population<sup>7,8</sup>. This is particularly relevant for tumours – cell populations in which rare phenotypes play a crucial role in the disease recurrence and outcome<sup>9,10</sup>. Continuous single cell imaging systems have provided important breakthroughs regarding the quality and throughput of images obtained from microscopy experiments. However, the sheer volume of data generated is still a challenge for most research teams, due to the lack of reliable, fast and universal analytical pipelines to extract information and reach valuable results from thousands of images<sup>11</sup>. To help counteract this, a wide range of analytical software is available to researchers, many stemming from scientific open source

communities<sup>12-17</sup>. Although they have different specific purposes, most software incorporate some sort of automation to alleviate the manual labour of detecting, counting and tracking individual cells in microscopy images.

Data analysis and visualization are two extremely relevant aspects of a successful single cell analysis pipeline. However, the availability of automated tools for such tasks is still limited, and the integration with existing cell detection software is a challenge for most users, normally requiring some degree of programming knowledge or leading to the development of ad-hoc, empirical solutions<sup>11</sup>. Here we introduce a Python-based tool called Family Tree Explorer (FATE) for cell tracking and lineage inference based on single cell detection in microscopic images (Fig. 1). Leveraging on the cell detection, measurement and segmentation performed by CellProfiler<sup>15</sup>, FATE is able to rebuild the ancestral history for each cell detected in a series of images. These histories, or “cell trees”, are able to retain and display information of every cell belonging to it. FATE was designed to help scientists by allowing easy visualization of cell trees, as well as analysis of the information contained in each tree; furthermore, it is also possible to use it to explore the heterogeneity of each cell inside a cell tree, or among cell trees themselves.

A typical FATE workflow is divided in 3 main steps: (i) data acquisition and preparation, (ii) FATE analysis, and (iii) output generation (Fig. 1). After acquiring a sequence of microscopy images, individual cells in each image are detected and measured using a custom CellProfiler pipeline (see Methods section for details). FATE uses the output from CellProfiler’s *ExportToSpreadsheet* module as its input; this module generates a spreadsheet-like text file containing information of each detected object, including its size, image in which it appears, XY position, and various area measurements.



**Figure 1. Overview of a typical FATE workflow.** CellProfiler's detection of cells in microscopy images is used as input for FATE. Ancestor trees are built in the program's memory while keeping track of each cell and its attributes. These trees can be programmatically explored and used to generate graphical and analytical output.

Each object detected becomes an object inside FATE's memory. All associated parameters (e.g. size, XY position, timepoint) are initialized as instance attributes. After all objects have been initialized, FATE assign cell tracks using an initial tracking algorithm. A cell track can be imagined as a line connecting a cell from the first to the final frame of a sequence of images. The tracking algorithm would be sufficient for cell tracking if the total number of cells did not change between timepoints; since cells migrate, divide and die, an algorithm to account for these events is necessary. In this step, the mother-daughters-sister relationship between cells is established. As mitosis regularly produces two cells, the algorithm was designed so that each cell can divide in two cells at most between two subsequent timeframes.

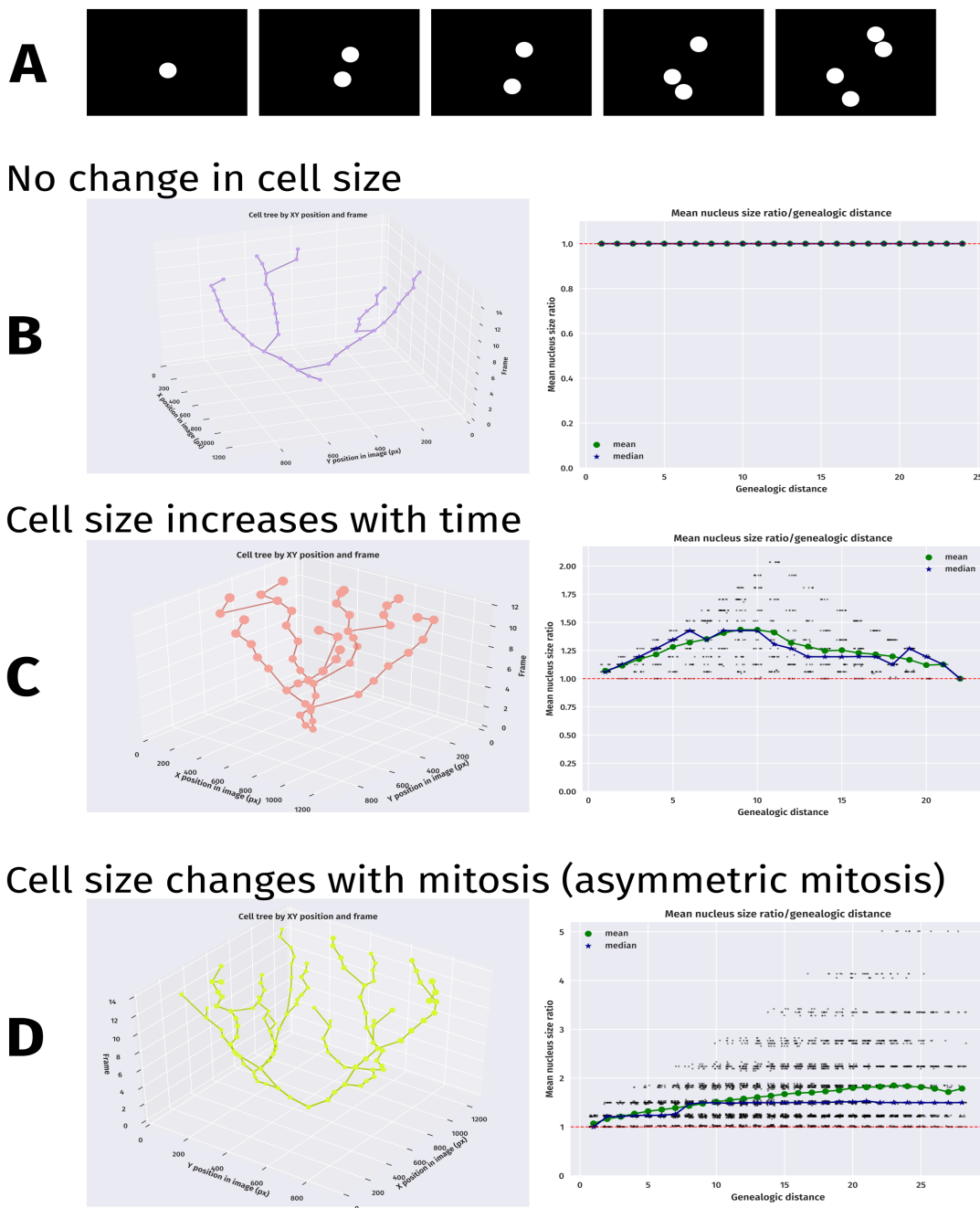
This model of mother-daughters-sister cells triplets resemble a binary tree structure (Fig. 1); indeed, the next step performed by FATE is to translate these inferred cell relationships in a more classical binary tree, with proper nodes and branches, along with computational methods for inserting new nodes, recursively walking down the tree, returning leaf nodes and children of a specific node, as well as returning the common ancestor of any two given nodes. This is a crucial step because it allows FATE to programmatically refer to the cell trees themselves, rather than just to "cells which belong to some unknown tree". This step also integrates the biological relevance of mother, daughter and sister cells migrating and dividing along time into a series of well-known and defined computational methods, simplifying the analytical process both to the user and the computer itself.

The last step is to analyse the trees and generate informative output. Since every cell in every tree carries information of its many characteristics as attributes, there is a wide array of possible output to be generated. One of the most visually informative output is to plot each tree in a 3D plot, with the XY axis corresponding to each cell's position and the Z axis representing the timepoint. The resulting plot shows the trees "growing" as time passes, and how each individual cell inside these trees have migrated. Another possibility is to plot any characteristic such as cell size, nucleus shape or fluorescence intensity along time; the resulting plot shows how the specified parameter changes as time passes, both for cells in any given cell tree, for trees themselves and for the population as a whole.

We have also explored a concept we named *genealogical distance* (gendist), which represents the distance between any two given cells inside the same cell tree. Any connection between cells, regardless of whether it represents a cell division or not, is representative of a

gendist of 1. For example, cells adjacent to one another inside the tree have a gendist of 1; sisters always have a gendist of 2; the gendist of a root cell and a leaf cell is equal to the length of the branch; and the gendist of any tree is highest when comparing cells in different branches, since it is necessary to walk down the branch from one cell and up the branch of the other to connect them. A possible analysis we explored is to plot the gendist of pairs of cells inside a same cell tree and compare it with some shared aspect of this cell pair, for example the ratio of each cell's nucleus size. When every cell is paired against every other cell in the cell tree, this analysis outputs a distribution particular to that specific cell tree.

In order to validate and regulate the analysis performed by FATE, we first created artificial datasets composed of simple white-on-black objects that could represent fluorescent cells in a experiment (Fig. 2A). We tested this approach with different test conditions for the objects: constant-sized objects (Fig. 2B), object size increasing with time (Fig. 2C), object size increasing with mitosis (not shown) and object size randomly changing with mitosis, mimicking asymmetric cell division (Fig. 2D)<sup>18</sup>. Cell tracking and tree construction worked as expected for all test cases, and graphical output was in accordance to the designed image sequences. Of particular interest, we detected very distinct aspects in the plots comparing the gendist with the size ratio of every possible pair of cells in each tree. For constant-sized objects, the size ratio is, unsurprisingly, also constant no matter what the genetic distance is (Fig. 2B). For objects whose size increases with time, however, the plot resembles a normal distribution, with cell pairs with similarly-sized nuclei associated with low and high values of gendist (Fig. 2C). This makes sense, since low values of gendist represent cells adjacent to each other, and high values of gendist are, in turn, comparisons between two cells in distant branches but at close timepoints – the main predictor of cell size in this example. Finally, objects mimicking asymmetric mitosis present a negatively skewed distribution (Fig. 2D).



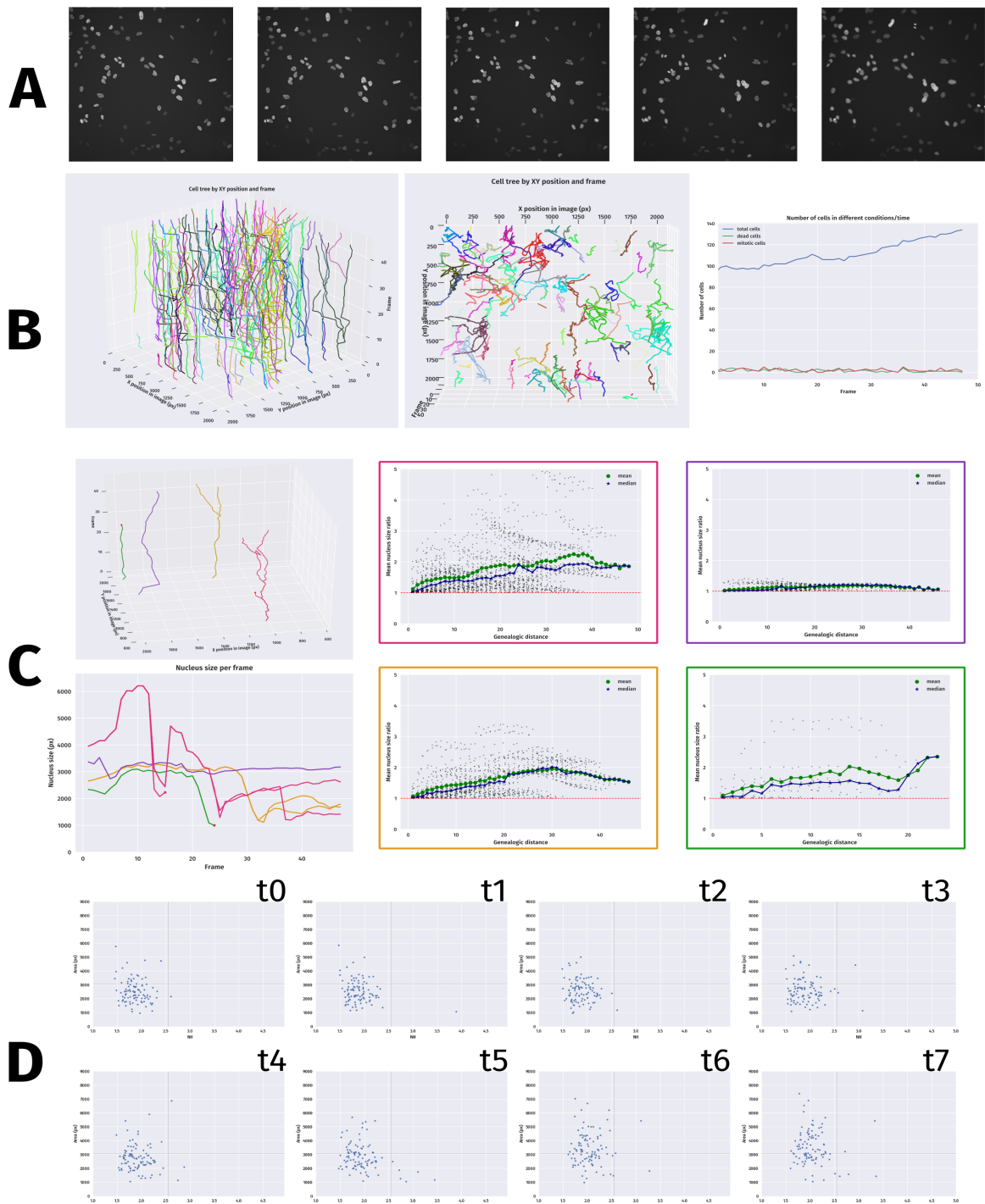
**Figure 2. Theoretical approaches of cell tree exploration.** **A**, example image sequence of simple white-on-black circular objects used on the theoretical approaches described here. **B**, cell tree in which cell size is constant. Since all cells have the same size at any time and in any branch, the nucleus size ratio remains constant for all genealogical distances. **C**, cell tree of cells growing in size as time passes. The distribution of nucleus size ratio per genealogical distance resembles a normal distribution. The nucleus size ratio increases with genealogical distance as leaf-to-root comparisons are made, but decreases as leaf cells are compared one to another, due to them being on the same time point. **D**, cell tree of cells with random fluctuations on cell size with each division. This mimics a situation in which mitosis is asymmetric. The distribution of nucleus size ratio per genealogical distance is negatively skewed, an indication that the shape of this distribution can possibly be used to infer intrinsic aspects of the cell tree and its history.

In order to assert the robustness of the analysis, we tested FATE using real data from fluorescence microscopy images. In a similar fashion to the previous analyses, nuclei from 45 sequential images (Fig. 3A) of A172-H2B-mCherry cell line were detected with CellProfiler and the output data was used as input by FATE. The program is able to scale the same graphical output of “growing trees” for multiple trees and timepoints (Fig. 3B, left). As previously discussed, FATE is also capable of generating visualizations for the total number of cells in each timepoint, and track nuclear size of each cell in each tree as time increases. We demonstrate that this capacity is true not only for our tests, but also for real image data (Fig. 3B right, Fig. 3C bottom-left). The output for gendist plots vs nucleus size ratio revealed a great variation in shape and trend for different cell trees (Fig. 3C).

As a proof of concept for further applicability and versatility of FATE, we also tested its capacity to automate the Nuclear Morphometric Analysis (NMA)<sup>19</sup>. Our preliminary results show that this is a possibility (Fig. 3D). In this manner, FATE has the ability to automate the conversion of CellProfiler’s output to the measurements required for a NMA, as described by Filippi-Chiela et al.

In this work, we described the conception and functionality of a novel tool for cell tracking, ancestor tree building and analytical plotting. FATE is able to correctly track cells and organize them in ancestor trees. Although software with similar goals exist<sup>16,20</sup>, we developed FATE not as a purely computational tool, but as a tool with biological questions at its core. In this manner, FATE is able to explore novel relations and reach new insights when analysing cells within a same cell tree, or when comparing different cell trees with one another.





**Figure 3. Analytical applications of FATE.** **A**, microscopy image sequence of A172-H2B-mCherry cells used as input for the described analyses. 400x magnification. 5 of a total of 45 frames are shown here. **B**, all detected cell trees are represented as trees (left), with their position on the XY axis changing as they grow over time (the Z axis). A top-down view (middle) reveals migration patterns. Total number of cells over time (right) is also a possible output. **C**, Selected cell trees (top-left) with different outcomes. FATE keeps track of cell size/nuclear size of each cell in each cell tree (bottom-left). Genealogical distance over time (middle, right) of each colour-coded cell tree reveals variation in the shape of the resulting distribution. **D**, Automation of the Nuclear Morphometric Analysis (NMA). After performing the necessary measurements for the NMA as previously described<sup>19</sup>, FATE is able to explore the relationship of nucleus size and nuclear irregularity for each cell in each timepoint in a continuous fashion.

The sheer variability of the gendist vs size ratio plots for the real image data was very surprising (Fig. 3B, right), especially since our test model displayed very characteristic plots for each scenario and condition (Fig. 2B – 2D). This leads us to conclude that the visual output of gendist vs cell size ratio could hold information about intrinsic aspects of each cell tree. This conclusion is not evident simply by looking at a cell tracking output, like those provided by other cell tracking software. “Whole tree” analyses are only possible due to FATE keeping track of many attributes for each single cell in each cell tree. We believe that these analytical standpoints, in which all cells inside a cell tree can be compared with other cells or group of cells, can potentially uncover novel information not yet explored by similar software. However, further exploration of these plots and of the gendist concept itself are necessary to confirm and expand this hypothesis.

There is still much room for improvement in FATE. We expect functionalities and output capabilities to improve as new hypotheses are made and new biologically-driven questions arise from its analyses. Even so, we believe that the current version of FATE is able to help scientists overcome the challenges of automating analyses of continuous single cell imaging and generating data in a fast and reliable fashion.

## **Methods**

### **Experimental setup**

*Cell culture conditions.* A172-H2B-mCherry cells were kindly obtained from MSc. Luana Lenz. Cells were cultured in DMEM culture medium (Gibco) supplemented with 10% FBS (Gibco) and no antibiotics. Cells were sub-cultured into new cell culture flasks once confluence reached 80%.

*Image acquisition.* Cells were seeded in a 96 well plate at a density of  $10^4$  cells per well and kept in DMEM medium with 10% FBS. 24 h after seeding, the plate was taken to a IN Cell Analyzer (GE Healthcare). This equipment includes a CO<sub>2</sub> chamber and plate heating, so that cells were kept at optimal culture conditions for the duration of the experiment. Cells were imaged every 30 minutes in the equipment, for a total of 45 images, using Cy3 filter for nuclei detection.

## **Image analysis**

*CellProfiler pipeline.* Images obtained from the previous step were analysed using a custom CellProfiler pipeline. The pipeline included the following modules: *Images*, for selecting input images; *Metadata*, for extraction of metadata from the file name of each picture; *NamesAndTypes* and *Groups*, which are required modules for CellProfiler initialization; *CorrectIlluminationCalculate* and *CorrectIlluminationApply*, to calculate and apply illumination correction to all input images; *IdentifyPrimaryObjects*, to detect cell nuclei as a single segmented object; *MeasureObjectIntensity* and *MeasureObjectSizeShape*, to make necessary measurements of each detected object; *ConvertObjectsToImage* and *SaveImages*, in order to save detected objects as images (used to assert object detection quality); and *ExportToSpreadsheet*, used to create the output text file used as input by FATE.

## **FATE analysis**

*Cell initialization and tracking.* Each line in the output file generated by CellProfiler is instantiated as a cell object in FATE's memory, with all associated parameters (e.g. size, XY position, timepoint) initialized as instance attributes. The exact parameters will depend on the analysis performed by CellProfiler. After object instantiation, FATE assigns cell tracks based on the following algorithm: start in the first timepoint  $t$ ; select one cell in this timepoint ( $c$ ); find cell  $d$  in timepoint  $t+1$  whose XY position is closest to  $c$ 's XY position; assign  $d$  as being the same cell as  $c$ , but in  $t+1$ ; do this for all cells in  $t$ ; then do this for all cells in all subsequent timepoints. Any cell in timepoint  $t+1$  already assigned to a cell in  $t$  cannot be reassigned and is ignored, forcing the algorithm to search for the next closest and available cell. At the end of the iteration, every cell  $c$  in any given timepoint  $t$  will be associated with a cell  $d$  in timepoint  $t+1$ , given that there was a least one cell  $d$  available for this association.

*Mother-daughters-sister relationship.* After cell initialization and tracking, the following algorithm is performed: Start with an unassigned cell  $c$  in a given timepoint  $t$ ; find the closest cell  $s$  also in  $t$  that is associated with a cell  $m$  in  $t-1$ ; reassign the three cells  $c$ ,  $s$  and  $m$  as follows:  $m$  is now the mother cell of  $c$  and  $s$ , appearing in  $t-1$ ;  $c$  and  $s$  are sister cells, appearing in  $t$ ;  $c$  and  $s$  are daughter cells of  $m$ ; repeat this for all unassigned cells in  $t$ ; then do this for all cells in all subsequent timepoints. If a possible sister cell  $s$  already has a sister and mother assigned to it, it is ignored, forcing the algorithm to search for cells with no mothers assigned to them.

*Tree representation.* All instantiated cells are reinstated as nodes of a binary tree. Cells with daughters are considered to have both a left and right branch; cells with a “next cell” (i.e. a cell  $d$  in  $t+1$  assigned as being the same as the cell  $c$  in  $t$ ) are considered as having a right branch only. Nodes are then organized in a Tree class, which acts as an ordered container for all nodes belonging to the same binary tree. Each binary tree has a direct correspondence to an ancestor tree inferred by FATE.

*Graphical output.* Plotting is done using the Matplotlib Python package<sup>21</sup>. All data is plotted through the standard Matplotlib API. For each plot, vectors of the values of interest (e.g. nucleus size, timepoint) are built by iteration and subsetting of all cells in memory. In particular, gendist calculations are performed by a Node method that calculates and returns the genealogical distance between any two cells in the same tree. By default, all possible plots are created and displayed; in order to simplify the user experience, a graphical user interface is currently under development.

## References

1. McGranahan, N. & Swanton, C. *Cell* **168**, 613–628 (2017).
2. Anderson, A. R. A., Weaver, A. M., Cummings, P. T. & Quaranta, V. Tumor. *Cell* **127**, 905–915 (2006).
3. Greaves, M. *Cancer Discov.* **5**, 806–820 (2015).
4. Hanahan, D. & Weinberg, R. A. *Cell* **144**, 646–674 (2011).
5. Zhao, B., Hemann, M. T. & Lauffenburger, D. A. *Trends Cancer* **2**, 144–158 (2016).
6. Landau, D. A., Carter, S. L., Getz, G. & Wu, C. J. *Leukemia* **28**, 34–43 (2014).
7. Altschuler, S. J. & Wu, L. F. *Cell* **141**, 559–563 (2010).
8. Wang, D. & Bodovitz, S. *Trends Biotechnol.* **28**, 281–290 (2010).
9. Murugaesu, N., Chew, S. K. & Swanton, C. *Am. J. Pathol.* **182**, 1962–1971 (2013).
10. Lipinski, K. A. *et al.* *Trends Cancer* **2**, 49–63 (2016).
11. Skylaki, S., Hilsenbeck, O. & Schroeder, T. *Nat. Biotechnol.* **34**, 1137–1144 (2016).
12. Eliceiri, K. W. *et al.* *Nat. Methods* **9**, 697–710 (2012).
13. Schindelin, J. *et al.* *Nat. Methods* **9**, 676–682 (2012).

14. Chaumont, F. de *et al. Nat. Methods* **9**, 690–696 (2012).
15. Carpenter, A. E. *et al. Genome Biol.* **7**, R100 (2006).
16. Chalfoun, J. *et al. Sci. Rep.* **6**, 36984 (2016).
17. Georgescu, W., Wikswo, J. P. & Quaranta, V. *Bioinformatics* **28**, 138–139 (2012).
18. Clevers, H. *Nat. Genet.* **37**, 1027–1028 (2005).
19. Filippi-Chiela, E. C. *et al. PLoS ONE* **7**, e42522 (2012).
20. Hilsenbeck, O. *et al. Nat. Biotechnol.* **34**, 703–706 (2016).
21. Caswell, T. A. *et al. Matplotlib/Matplotlib V3.0.2.* (2018).  
doi:10.5281/zenodo.1482099

## 5 PERSPECTIVAS FUTURAS

Este trabalho apresentou FATE, um programa capaz de rastrear células a partir de dados gerados pela detecção do programa aberto CellProfiler e agrupá-las em árvores de descendentes, as quais podem ser exploradas em função das células que as compõem e de características de cada célula.

Há ainda diversas melhorias a ser realizadas ao FATE, como por exemplo o desenvolvimento de uma interface gráfica para simplificar o uso, ou o aprimoramento dos algoritmos aplicados pelo programa para rastrear células e construir árvores. Além disso, são necessários mais testes para validar o uso do FATE para NMA, desenhando experimentos com condições de tratamento e modificando o *output* gerado para suportar diversas condições de cultivo organizadas em um mesmo gráfico.

Mesmo assim, a ferramenta em sua versão atual apresenta uma boa capacidade e diversidade analítica. Ela pode ser utilizada para quantificar automaticamente o crescimento de colônias celulares ao longo do tempo, ou para avaliar como algum aspecto de interesse – a intensidade de algum marcador fluorescente, por exemplo – flutua ao longo do tempo em cada célula, em cada árvore de descendentes e na população em geral na sequência de imagens. Acreditamos, contudo, que o diferencial da ferramenta está em seu *design* voltado para questões biológicas. As análises realizadas para células dentro de uma mesma árvore só são possíveis porque o programa guarda em memória o máximo de informação para cada célula inicializada.

Esperamos que com o desenvolvimento contínuo do FATE seja possível realizar novas análises similares a de distância genealógica vs razão de tamanho de núcleo, onde aspectos gerais de uma árvore de descendentes são gerados a partir de comparações individuais de seus componentes. Este tipo de análise pode levar ao surgimento de padrões até então desconhecidos em árvores de descendentes, à formulação de hipóteses biológicas e à construção de modelos preditivos de desfecho celular.

Em última análise, o crescimento de colônias celulares é um evento cotidiano, quase trivial. Contudo, observar aspectos comuns em biologia sob uma óptica inovadora requer um esforço criativo constante, e pode, por vezes, levar a resultados surpreendentes. Neste sentido, FATE auxilia os cientistas tanto a automatizar a análise de dados de células únicas em

imagens de microscopia, quanto a ampliar sua caixa de ferramentas analíticas em busca de novas descobertas e *insights* científicos.

## 6 REFERÊNCIAS DO TEXTO PRINCIPAL

Altschuler, S.J., and Wu, L.F. (2010). Cellular heterogeneity: when do differences make a difference? *Cell* 141, 559–563.

Balaban, N.Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004). Bacterial Persistence as a Phenotypic Switch. *Science* 305, 1622–1625.

Banks, V.A., Stanton, N.A., and Harvey, C. (2014). Sub-systems on the road to vehicle automation: Hands and feet free but not ‘mind’ free driving. *Saf. Sci.* 62, 505–514.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* 7, R100.

Chatterjee, M., and Acar, M. (2018). Heritable stress response dynamics revealed by single-cell genealogy. *Sci. Adv.* 4, e1701775.

Chaumont, F. de, Dallongeville, S., Chenouard, N., Hervé, N., Pop, S., Provoost, T., Meas-Yedid, V., Pankajakshan, P., Lecomte, T., Montagner, Y.L., et al. (2012). Icy: an open bioimage informatics platform for extended reproducible research. *Nat. Methods* 9, 690–696.

Collins, K.A., Kielkopf, J.F., Stassun, K.G., and Hessman, F.V. (2017). AstroImageJ: Image Processing and Photometric Extraction for Ultra-precise Astronomical Light Curves. *Astron. J.* 153, 77.

Correa-Baena, J.-P., Hippalgaonkar, K., van Duren, J., Jaffer, S., Chandrasekhar, V.R., Stevanovic, V., Wadia, C., Guha, S., and Buonassisi, T. (2018). Accelerating Materials Development via Automation, Machine Learning, and High-Performance Computing. *Joule* 2, 1410–1420.

Ding, L., Raphael, B.J., Chen, F., and Wendl, M.C. (2013). Advances for studying clonal evolution in cancer. *Cancer Lett.* 340, 212–219.

Eliceiri, K.W., Berthold, M.R., Goldberg, I.G., Ibáñez, L., Manjunath, B.S., Martone, M.E., Murphy, R.F., Peng, H., Plant, A.L., Roysam, B., et al. (2012). Biological Imaging Software Tools. *Nat. Methods* 9, 697–710.

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118.

Filippi-Chiela, E.C., Oliveira, M.M., Jurkovski, B., Callegari-Jacques, S.M., Silva, V.D. da, and Lenz, G. (2012). Nuclear Morphometric Analysis (NMA): Screening of Senescence, Apoptosis and Nuclear Irregularities. *PLOS ONE* 7, e42522.



- Hart, E.M., Barmby, P., LeBauer, D., Michonneau, F., Mount, S., Mulrooney, P., Poisot, T., Woo, K.H., Zimmerman, N.B., and Hollister, J.W. (2016). Ten Simple Rules for Digital Data Storage. *PLoS Comput. Biol.* *12*.
- Hawker, C.D. (2007). Laboratory Automation: Total and Subtotal. *Clin. Lab. Med.* *27*, 749–770.
- Hirsch, F.R., Ottesen, G., Pødenphant, J., and Olsen, J. (1983). Tumor heterogeneity in lung cancer based on light microscopic features. *Virchows Arch. A* *402*, 147–153.
- Hundeshagen, P., Hamacher-Brady, A., Eils, R., and Brady, N.R. (2011). Concurrent detection of autolysosome formation and lysosomal degradation by flow cytometry in a high-content screen for inducers of autophagy. *BMC Biol.* *9*, 38.
- Katajisto, P., Döhla, J., Chaffer, C.L., Pentimikko, N., Marjanovic, N., Iqbal, S., Zoncu, R., Chen, W., Weinberg, R.A., and Sabatini, D.M. (2015). Asymmetric apportioning of aged mitochondria between daughter cells is required for stemness. *Science* *348*, 340–343.
- Kester, L., and Oudenaarden, A. van (2018). Single-Cell Transcriptomics Meets Lineage Tracing. *Cell Stem Cell* *23*, 166–179.
- King, R.D., Rowland, J., Oliver, S.G., Young, M., Aubrey, W., Byrne, E., Liakata, M., Markham, M., Pir, P., Soldatova, L.N., et al. (2009). The Automation of Science. *Science* *324*, 85–89.
- Konczal, J., and Gray, C.H. (2017). Streamlining workflow and automation to accelerate laboratory scale protein production. *Protein Expr. Purif.* *133*, 160–169.
- Landau, D.A., Carter, S.L., Getz, G., and Wu, C.J. (2014). Clonal evolution in hematologic malignancies and therapeutic implications. *Leukemia* *28*, 34–43.
- Lechler, T., and Fuchs, E. (2005). Asymmetric cell divisions promote stratification and differentiation of mammalian skin. *Nature* *437*, 275–280.
- Luyi, T., Schreuder, J., Amann-Zalcenstein, D., Tran, J., Kocovski, N., Su, S., Diakumis, P., Bahlo, M., Sargeant, T., Ritchie, M., et al. (2018). SIS-seq, a molecular “time machine”, connects single cell fate with gene programs. *BioRxiv* 403113.
- Mattiazzi Usaj, M., Styles, E.B., Verster, A.J., Friesen, H., Boone, C., and Andrews, B.J. (2016). High-Content Screening for Quantitative Cell Biology. *Trends Cell Biol.* *26*, 598–611.
- McGranahan, N., and Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* *168*, 613–628.
- Morrison, S.J., and Kimble, J. (2006b). Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature* *441*, 1068.

- Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D., et al. (2011). Tumor Evolution Inferred by Single Cell Sequencing. *Nature* 472, 90–94.
- Nawy, T. (2014). Single-cell sequencing. *Nat. Methods* 11, 18–18.
- Novak, S.M., and Marlowe, E.M. (2013). Automation in the Clinical Microbiology Laboratory. *Clin. Lab. Med.* 33, 567–588.
- Ornatsky, O., Bandura, D., Baranov, V., Nitz, M., Winnik, M.A., and Tanner, S. (2010). Highly multiparametric analysis by mass cytometry. *J. Immunol. Methods* 361, 1–20.
- Riben, M. (2015). Laboratory Automation and Middleware. *Surg. Pathol. Clin.* 8, 175–186.
- Rotman, D. (2017). The Relentless Pace of Automation.
- Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., Kashiwagi, S., Fukami, K., Miyata, T., Miyoshi, H., et al. (2008). Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell* 132, 487–498.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682.
- Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis.
- Sirachý, J. (1979). An approach to the problem of heterogeneity of human tumour-cell populations. *Br. J. Cancer* 39, 570–577.
- Skylaki, S., Hilsenbeck, O., and Schroeder, T. (2016). Challenges in long-term imaging and quantification of single-cell dynamics. *Nat. Biotechnol.* 34, 1137–1144.
- Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M., and Sorger, P.K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 459, 428–432.
- Spiller, D.G., Wood, C.D., Rand, D.A., and White, M.R.H. (2010). Measurement of single-cell dynamics. *Nature* 465, 736–745.
- Strome, S., and Wood, W.B. (1983). Generation of asymmetry and segregation of germ-line granules in early *C. elegans* embryos. *Cell* 35, 15–25.
- Tellez-Gabriel, M., Ory, B., Lamoureux, F., Heymann, M.-F., and Heymann, D. (2016). Tumour Heterogeneity: The Key Advantages of Single-Cell Analysis. *Int. J. Mol. Sci.* 17.
- Wang, D., and Bodovitz, S. (2010). Single cell analysis: the new frontier in ‘Omics.’ *Trends Biotechnol.* 28, 281–290.

Wills, Q.F., and Mead, A.J. (2015). Application of single-cell genomics in cancer: promise and challenges. *Hum. Mol. Genet.* *24*, R74–R84.

Woodruff, M.F. (1983). Cellular heterogeneity in tumours. *Br. J. Cancer* *47*, 589–594.

Yamashita, Y.M., Jones, D.L., and Fuller, M.T. (2003). Orientation of Asymmetric Stem Cell Division by the APC Tumor Suppressor and Centrosome. *Science* *301*, 1547–1550.

Zavadskas, E.K. (2010). Automation and robotics in construction: International research and achievements. *Autom. Constr.* *19*, 286–290.

## 7 ANEXOS

### 7.1 Normas de publicação para Nature Methods – Brief Communication

Normas disponíveis em: <https://www.nature.com/nmeth/about/content>.

A Brief Communication is a more concise format used typically to report a significant improvement to a tried-and-tested method, its modification and adaptation to an important original application, or an important new tool or resource of broad interest for the scientific community. This format typically does not exceed 3 printed pages. Brief Communications begin with a brief unreferenced abstract (3 sentences, no more than 70 words), which will appear on Medline. The title is limited to 10 words (or 90 characters). The main text is typically 1,000-1,500 words, including the abstract and contains no headings with the exception of a single heading for Methods to point readers to the online Methods section providing all technical details necessary for the independent reproduction of the methodology. Brief Communications normally have no more than 2 display items, although this may be flexible at the discretion of the editor, provided the page limit is observed. As a guideline, Brief Communications allow up to 20 references, and article titles are omitted from the reference list.

Brief Communications include received/accepted dates. They may be accompanied by supplementary information. Brief Communications are peer reviewed.