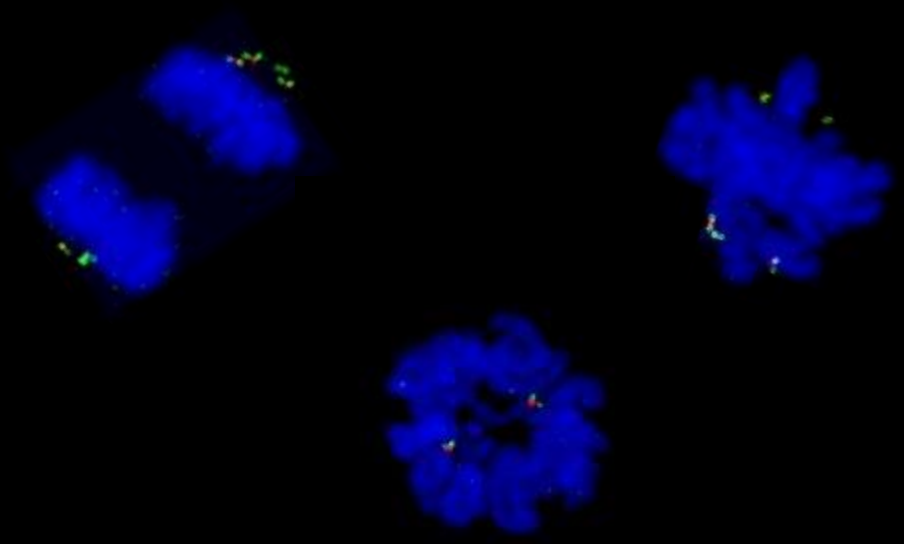
2024

Population dynamics of centrosome amplification

Marco António Dias Louro



Dissertation presented to obtain the **Ph.D degree in**

Integrative Biology and Biomedicine

Oeiras, Maio 2024

INSTITUTO DE TECNOLOGIA QUÍMICA E BIOLÓGICA ANTÓNIO XAVIER | UNIVERSIDADE NOVA DE LISBOA

ITQB NOVA

DOCTORAL THESIS

Population dynamics of centrosome amplification

Author:
Marco António DIAS LOURO

Supervisors:
Dr. Mónica BETTENCOURT-DIAS
Dr. Claudia BANK

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Cell Cycle Regulation & Evolutionary Dynamics Groups
Integrative Biology and Biomedicine PhD Programme
Instituto Gulbenkian de Ciência

Declaration

I, Marco António DIAS LOURO, declare that this thesis titled "Population dynamics of centrosome amplification" and the work presented in it are my own. In particular, I declare that:

- This work was done while in candidature for a research degree at this University.
- Where I have consulted or quoted from the published work of others, this is always clearly attributed.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.
- I have followed the rules of scientific good practice.

This work was funded by the Fundação Calouste Gulbenkian and the Fundação para a Ciência e a Tecnologia (FCT) PhD funding grant PD/BD/139217/2018 awarded to the student. Additional funding was provided by the EMBO Installation Grant IG4152, ERC Starting Grant 804569 – FIT2GO, and FCT grant PTDC/BIA-BID/32225/2017, awarded to Claudia Bank, the ERC Consolidator Grant 683258 - CentrioleBirthDeat, awarded to Mónica Bettencourt-Dias, and the University of Bern.

This work was supported by the thesis committee composed of Dr. Jorge Carneiro and Dr. Lars Jansen, and will be defended in front of the following examiners: Dr. Renata Basto, Dr. Andrew Murray, Dr. Francisco Dionísio, and Dr. Élio Sucena.

ITQB NOVA

Abstract

Instituto Gulbenkian de Ciência

Integrative Biology and Biomedicine PhD Programme

Doctor of Philosophy

Population dynamics of centrosome amplification

by Marco António DIAS LOURO

Living organisms are fundamentally composed of cells and subject to evolutionary change. These are perhaps two of the most general concepts in biology. An integrative framework that encompasses both evolutionary and cell biology is necessary for understanding how cellular features change and diversify. Such a framework is crucial for understanding a plethora of phenomena, ranging from how cellular components expand and disappear across the tree of life to human disease. In this thesis, I make use of evolutionary and cell biological principles to address the causes and consequences of centriole number abnormalities.

Centrioles are the core structure of cilia and the eukaryotic centrosome. As centrosomal components, centrioles play a key role in organising the microtubule cytoskeleton and thus, are central to a multitude of subcellular processes, such as the establishment of cell polarity and chromosome segregation during mitosis. Typical healthy cells contain two centrioles, which are duplicated once and only once per cell cycle. This is crucial for ensuring genome integrity and cell viability. Anomalous centriole

numbers induce a variety of cellular defects and have been implicated in human disease, most notably in cancer.

Cancer cells usually display abnormally high numbers of centrioles, or centrosome amplification. These numbers tend to be very heterogeneous within and between cancer cell populations, but their distribution appears to be stable. It is thought that negative selection of cells with extra centrioles balances out errors in centriole biogenesis, thus leading to an equilibrium. However, how these dynamics are generated and their consequences for cellular physiology were poorly understood.

In chapter II, we developed mathematical models inspired by mutation-selection processes for inferring patterns of centriole overproduction and negative selection in a broad panel of cancer cell lines. Our results indicated that multi-step overproduction events as well as constant selection against cells with any number of extra centrioles can explain most empirical distributions of centriole numbers per cell.

In chapter III, we propose an extended model that explicitly accounts for how extra centrioles are segregated at mitosis and how centriole numbers change along the cell cycle. We provide computational support for the notion that asymmetric segregation of centrioles can allow cells to tolerate centrosome amplification and could potentiate cancer development for modest rates of centriole overproduction. Our models further predict that the distribution of centriole numbers per cell is sensitive to when selection acts, which can lead to different degrees of underscoring of centriole number anomalies in interphase cell populations.

In chapter IV, we devise a long-term cell culturing setup for assessing centriole number and fitness dynamics in human cell lines chronically overexpressing a key regulator of centriole biogenesis. We found that chronic overexpression of this protein at different levels did allow for sustained centrosome amplification.

We further report that centrosome amplification imparts a proliferation deficit that is proportional to the relative frequency of cells with extra centrioles.

Taken together, our work yielded novel insights on the processes that govern the dynamics of anomalous centriole numbers and how they may affect cancer development.

Keywords: centrosome amplification, mutation-selection balance, cancer, Plk4, cell cycle

ITQB NOVA

Resumo

Instituto Gulbenkian de Ciência

Biologia Integrativa e Biomedicina

Grau de Doutor

Dinâmica populacional da amplificação de centrossomas

por Marco António DIAS LOURO

Os organismos vivos são fundamentalmente compostos por células e estão sujeitos a alterações evolutivas. Estes conceitos são dois dos mais gerais em biologia. Uma perspetiva integrativa da biologia evolutiva e da biologia celular é necessária para compreender como determinadas características celulares se modificam e diversificam. Esta perspetiva é crucial para compreender uma variedade de fenómenos, desde a expansão ou o desaparecimento de componentes celulares ao longo da árvore da vida, à doença humana. Nesta tese, utilizo princípios da biologia evolutiva e da biologia celular para abordar as causas e consequências de anomalias no número de centríolos.

Os centríolos são as estruturas principais dos cílios e do centrossoma eucariótico. Enquanto componentes centrossomais, os centríolos têm um papel-chave na organização do citosqueleto de microtúbulos e, portanto, são centrais numa multitude de processos subcelulares, como o estabelecimento de polaridade celular e a segregação de cromossomas durante a mitose. Tipicamente, as células saudáveis contêm dois centríolos, que são duplicados uma e só uma vez a cada ciclo celular. Isto é crucial para garantir

a integridade do genoma e viabilidade das células. Números anómalos de centríolos induzem uma variedade de defeitos celulares e têm implicações em doenças humanas, nomeadamente, no cancro.

As células cancerígenas normalmente possuem um número anormalmente alto de centríolos, denominado amplificação de centrossomas. Estes números tendem a ser extremamente heterogéneos dentro de e entre populações de células cancerígenas, mas a sua distribuição na população tende a ser estável. Pensa-se que a seleção negativa de células com centríolos extra equilibre erros na biogénese de centríolos, levando a um estado de equilíbrio. Contudo, sabe-se pouco como esta dinâmica é gerada e quais as suas consequências para a fisiologia das células.

No capítulo II, desenvolvemos modelos matemáticos inspirados em processos de mutação-seleção para inferir padrões de sobre-produção de centríolos e seleção negativa num painel alargado de linhas celular de cancro. Os nossos resultados indicam que eventos de sobre-produção em múltiplos passos e uma pressão selectiva constante contra células com centríolos extra, independentemente do número, explicam a maioria das distribuições empíricas de números de centríolos por célula.

No capítulo III, propomos um modelo mais extenso que inclui explicitamente segregação de centríolos extra durante a mitose e alterações do número de centríolos ao longo do ciclo celular. Fornecemos suporte computacional para a noção de que a segregação asimétrica de centríolos pode permitir que as células tolerem a amplificação de centrossomas e poderá potenciar o desenvolvimento de cancro para taxas modestas de sobre-produção de centríolos. Os nossos modelos preveem também que a distribuição do número de centríolos por célula é sensível ao momento em que a seleção actua, o que pode levar a diferentes graus de subestimação de anomalias no número de centríolos em populações de células interfásicas.

No capítulo IV, delineámos um sistema de cultura de células a longo prazo para avaliar dinâmicas de número de centríolos e "fitness" em linhas celulares humanas a sobre-expressar cronicamente um regulador-chave da biogénese de centríolos, a Plk4. Descobrimos que diferentes níveis de sobre-expressão crónica desta proteína não permitem a amplificação sustentada de centrossomas. Reportamos também que a amplificação de centrossomas induz um défice proliferativo que é proporcional à frequência relativa de células com centríolos extra.

No cômputo geral, o nosso trabalho gerou novas perspectivas sobre os processos que governam a dinâmica de números anómalos de centríolos e como os mesmos podem afetar o desenvolvimento do cancro.

Palavras-chave: amplificação de centrossomas, equilíbrio mutação-seleção, cancro, Plk4, ciclo celular

Errata

Chapter II

Page 74 - Fig. 6 legend "Black dots indicate the median estimated value and the green dot indicates the input value used in the simulations."

The legend should read: "Black dots indicate the input value used in the simulations and red squares indicate the median estimated value."

Chapter IV

In this chapter, we chronically overexpressed PLK4, a key regulator or centriole biogenesis, in the breast cell line MCF10A-PLK4. Initially after PLK4 overexpression, MCF10A-PLK4 acquired extra centrioles in a dose-dependent manner. However, we observed a reduction in the relative frequency of cells with extra centrioles over time, approaching basal levels after two months.

These evolutionary dynamics of MCF10A-PLK4 could potentially be explained by cells bearing centriole amplification being outcompeted by a mutant already present in the ancestral population that could not generate extra centrioles, or by the appearance of a mutant during the evolution experiment. We addressed this hypothesis by performing bulk RNA-seq of cells sampled at days 1, 3, 31, and 64 (final time point) of experimental evolution. We started by performing variant analysis of PLK4 that could inhibit centriole overproduction. If this hypothesis was correct, we expected that such a variant would be observed at low frequencies or not detected in the ancestral population, and rise in frequency over time.

As expected, we did not detect variants displaying such a pattern in MCF10A-TetR and MCF10A-PLK4$^{1-608}$. For MCF10A-PLK4, we detected reads containing a SNP which results in a stop codon yielding the truncation of the exogenous PLK4 present

in MCF10A-PLK4$^{1-608}$. We detected this SNP at days 31 and day 64 in both evolution experiments for the populations treated with Dox 0.1., being higher in the latter time point. The SNP was not detected in MCF10A-PLK4 evolved in Dox 0- or Dox0.001. At this time we propose that the ancestral MCF10A-PLK4 population was contaminated with a small proportion of MCF10A-PLK4$^{1-608}$ and that with higher Dox concentration imposing a stronger selective pressure against cells with centrosome amplification, the cells overexpressing the truncated version of Plk4 became more abundant over time.

It is possible that in cancer evolution different variants of PLK4 are expressed and compete as truncated forms were recently identified in cancer[1]. Thus, our results may be relevant for shedding light on the evolutionary dynamics of cell populations with and without centrosome amplification, in particular during cancer development. We are currently quantifying the relative frequency of MCF10A-PLK4$^{1-608}$ cells in MCF10A-PLK4 along experimental evolution and will update the data and its interpretation in this chapter accordingly.

An updated version can be found in:
https://www.biorxiv.org/content/10.1101/2024.02.15.580424v2

References

1. Kasera, H., Singh, P. PLK4 homodimerization is required for CEP152 centrosome localization and spindle organization. *bioRxiv* 2023.11.06.565834; doi: https//doi.org/10.1101/2023.11.06.565834

Contents

Declaration iii

Abstract v

Resumo ix

Errata xiii

1 Introduction 1

2 Patterns of selection against centrosome amplification 61

3 A cell cycle-structured model of centriole number dynamics 89

4 Short-term evolution of cells with centrosome amplification 143

5 Discussion 207

Acknowledgements 231

* References are included at the end of each chapter.

List of Abbreviations

ANKRD6	Ankyrin repeat domain 6
BRCA1	Breast cancer type 1 susceptibility protein
BRCA2	Breast cancer type 2 susceptibility protein
CA	Centrosome amplification
Cep97	Centrosomal protein of 97 KDa
Cep135/Bld10	Centrosomal protein of 135 KDa
Cep152	Centrosomal protein of 152 KDa
Cep192	Centrosomal protein of 192 KDa
CFSE	Carboxyfluorescein succinimidyl ester
CP110	Centriolar coiled-coil protein of 110 kDa
CPAP	Centrosomal P4.1-associated protein
DLD-1	Human colorectal adenocarcinoma cell line
DNA	Deoxyribonucleic acid
EGF	Epidermal growth factor
ESPL-1	Extra spindle pole bodies like 1, separase
IAA	Indole-3-acetic acid
MAPK	Mitogen-activated protein kinase
MCF10A	Human non-transformed mammary epithelium cell line
Mdm2	Murine double minute 2
mRNA	Messenger ribonucleic acid
ODF2	Outer dense fiber of sperm tails 2
PI	Propidium iodide
Plk1	Polo-like kinase 1
Plk4	Polo-like kinase 4
NFκB	Nuclear factor κ-light-chain-enhancer of activated B cells

p53	Tumour protein 53
qPCR	Quantitative polymerase chain reaction
RPE	Retinal pigment epithelium
RKO	Human colorectal cancer cell line
SAS6	Spindle assembly abnormal protein 6
STIL	SCL/Tal-1 interrupting locus
U2OS	*Homo sapiens* bone osteosarcoma

For ’salinda, Maria, and Carolina.
For "avô João" and "avó Linda".

Chapter 1

Introduction

Foundations and challenges of evolutionary cell biology

Evolution has long been concerned with explaining how diverse lifeforms come to be. Darwin and Wallace's theory of evolution through natural selection provided a powerful conceptual idea of how evolution could operate (Darwin, 1859). It was grounded on three core premises. First, natural selection requires naturally occurring variation in species traits. Second, some traits should lead to differential reproduction, i.e. individuals with a certain characteristic should generate more offspring than others. Third, advantageous traits should be heritable. Thus, individuals with advantageous heritable characteristics would pass them on to their descendants, which would become more abundant in the population. However, the original theory of natural selection was agnostic to the vehicle of heredity and how phenotypic variation could be produced, notwithstanding Darwin's hypothesis on the matter, which was later disproved (Darwin, 1859). The modern synthesis brought about in the beginning of the 20th century by the likes of Ronald Fisher (Fisher, 1922), Sewall Wright (Wright, 1938), J. B. S. Haldane (Haldane, 1927), reconciled and expanded upon Mendelian genetics and evolution by natural selection, marking the advent of population genetics. Following the discovery of DNA as the molecule of heredity, population genetics recognised mutations as the principal source of variation in natural populations. Incorporating other novel concepts such as allelic variation and genetic drift, population genetics founded a quantitative theory of evolution (Huxley et al., 2010). To this day, evolutionary theory has been hailed as the closest example in biology of a unified theory (Smocovitis, 1992). This idea was famously encapsulated in a quote by Theodosius Dobzhansky, "nothing in biology makes sense except in the light of evolution" (Dobzhansky, 1973).

Cell biology has been chiefly interested in describing the form, diversity, and function of cells and their components. The history of cell biology is closely linked to the invention of the microscope and pioneering observations from various researchers. The term "cell" was coined by Hooke when he observed that plants appeared to be composed of minuscule honeycomb-like compartments. Later, Anton van Leeuwenhoek provided the first description of microorganisms, which he dubbed "animalcules". The realisation that all living organisms seemed to be composed of cells formed the basis of the cell theory, championed by Schleiden and Schwann. As microscopes became increasingly refined, and novel chemical methods became available, the intracellular medium - the cytosol, originally named "protoplasm" - and various subcellular constituents, such as mitochondria and the endoplasmic reticulum were characterised (Mazzarello, 1999). The development of cell culture methods, genetic manipulation tools, electron and fluorescence microscopy techniques in the past century turned cell biology into a fully-fledged experimental science (Cooper, 2000). The focus of research shifted from observation and description of different cell types to engineering principles underlying cellular architecture. Cytoskeletal dynamics, biosynthesis of subcellular structures, regulation of gene expression and protein abundance, among others are central questions in the field. Nowadays, cell biology is progressing from being a largely empirical science to a markedly quantitative field, with important contributions from biophysics and introgression of novel, if controversial, concepts in our understanding of cells, such as the role of phase separation in cellular compartmentalisation (Boeynaems et al., 2018; Hyman et al., 2014).

The field of evolutionary cell biology aims at unifying cell and evolutionary biology in order to understand how cellular features evolve. Historically, it was grounded on comparative cell biology, as well as comparative genomics i.e. the comparison of

cells or cellular features across different organisms, seeking to describe how certain structures expanded or disappeared across the tree of life (Islas-Morales et al., 2021). More recently, there have been calls for a more comprehensive approach (Islas-Morales et al., 2021; Lynch et al., 2014; Lynch et al., 2022). Cell biology can, in principle, fuel evolutionary biology with detailed mechanistic insights on specific biological systems, whereas evolutionary biology can provide cell biology with a broader quantitative framework for explaining how these mechanisms arose and evolved (Lynch et al., 2022). Whereas evolutionary biology is grounded on a well-developed mathematical theory, it suffers from a high degree of abstraction and, potentially, over-simplification. Population genetics often disregards the intricacies of specific biological systems, which might limit its predictive power (Welch, 2017). Cell biology, on the other hand, lacks a foundational mathematical theory, despite some well established general principles (Islas-Morales et al., 2021). The limitations of both fields suggest a high degree of complementarity but a quantitative framework linking the two is still lacking.

Recently, a proposal centred on energy budget of subcellular structures such as organelles and cilia has been put forward as an attempt at a unified framework of evolutionary cell biology (Lynch et al., 2022). The core premise of this proposal lies in the fact that cells spend resources on building and maintaining such structures via biosynthetic processes. Since cells cannot access unlimited resources, investing them into biosynthesis deems that the cell is not allocating them for other activities, such as regulating osmotic pressure or pH. Therefore, the selective advantage conferred by possessing a certain structure must compensate its energy cost in order for said structure to be maintained or expand. A bioenergetics-centred approach has the potential to be universal since living organisms utilise similar energy sources (e.g. ATP) and it can provide a mechanistic baseline for understanding the

patterns of gains and loss of certain subcellular features across evolution (Lynch et al., 2022). However, the predictive power of this theory is yet to be tested.

Conversely, others have combined evolutionary principles with simple molecular notions to explain, for instance, the evolution of gene regulatory networks. For example, Friedlander *et al.* investigated the evolution of gene regulation following gene duplication of a given transcription factor (Friedlander et al., 2017). They developed a mathematical model taking into account the biophysics transcription factor binding depending on the number of mismatches between its consensus sequence and genomic binding site. Gene expression of target genes depended on the efficiency of binding and on two environmental signals, which could be correlated or anti-correlated in time. The expression level of said target genes determined fitness. This approach provided a more realistic scenario to study how each of the transcription factor gene copies evolved depending on the selective pressure on gene expression and the degree of correlation between the two environmental inputs.

Delving on the same topic, Lynch (Lynch, 2007) provided a neutral expectation for the ratio of gains versus losses of transcription factor binding sites as a function of the population size, mutation rate, and genome size. Making use of available estimates for these parameters in a variety of prokaryotic and eukaryotic lineages, the author concluded that gains of transcription factor binding sites should be favoured in groups with larger genomes, such as vertebrates, whereas the gene regulatory networks of prokaryotes should be more streamlined. The key insight in this study is that complexity in gene regulation networks or lack thereof can be explained without invoking the action of selection. This contrasts with the near-ubiquitous assumption that adaptive processes drove the evolution of complexity. Therefore, a population genetics approach can offer new perspectives on

classical problems in biology.

Overall, evolutionary biology and cell biology are two disciplines are becoming increasingly intertwined. Although a consensual unifying framework is still lacking, the examples above serve to demonstrate how the molecular inner workings of a cell can constrain evolution, and how evolutionary biology can offer novel perspectives on cell biological problems.

Thesis outline

In this thesis, we apply classical concepts of population genetics to the problem of copy number regulation of subcellular structures. Our subject of focus is the centriole, a protein complex present in several eukaryotes. Centriole number in typical cycling cells displays dynamics similar to those of DNA replication (Brito et al., 2012) - each of the two centrioles is precisely duplicated and each daughter cell inherits a pair of centrioles after mitosis. Whereas centriole number is almost invariant in healthy tissues, it is widely variable in cancer (Lopes et al., 2018; Marteil et al., 2018, reviewed in Chan, 2011; Godinho and Pellman, 2014). However, how this heterogeneity is generated and the role it plays in cellular physiology is poorly understood.

In the following sections I briefly describe centriolar structure and the canonical duplication/segregation cycle. Then, I explore the impact of centriole numeric anomalies in cancer development. Next, we review how errors in centriole duplication are generated and counter-selected, and their resulting temporal dynamics. Based on these studies, we propose that centriole number dynamics can be likened to a mutation-selection process and that the presence of extra centrioles can act as a mutator phenotype. I conclude the chapter by presenting an overview of the research work conducted for this thesis, and how we used mathematical

modelling and experimental evolution to study the causes and consequences of centriole number alterations over time.

The centriole duplication and segregation cycle

Centrioles are diminutive protein-based structures that occur in a variety of eukaryotic lineages. They are generally composed of nine microtubule triplets arranged radially, although variations to this organisation are found across the tree of life (reviewed in Jana, 2020; Nabais et al., 2018; Nabais et al., 2020). Centrioles have two main roles inside the cell. First, centrioles can mature into basal bodies, laying the groundwork for cilia assembly. Second, centrioles can organise the centrosome by recruiting an intricate matrix of proteins called the pericentriolar material (PCM). This thesis is particularly focused on centrioles as core structures of the centrosome. The centrosome is the main microtubule-organising centre (MTOC) in cycling vertebrate cells. As the terminology implies, it is responsible for creating a favourable environment for microtubule nucleation. Thus, centrosomes are especially relevant for any cellular process involving the microtubule cytoskeleton, such as the establishment of cell polarity, signalling, and the formation of the mitotic spindle (reviewed in Breslow and Holland, 2019; Brito et al., 2012; Gomes Pereira et al., 2021; Gönczy, 2015).

In many types of proliferating cells, the G1 centrosome contains two centrioles which are duplicated through a self-organising process starting at the G1-S transition (reviewed in Gomes Pereira et al., 2021; Karsenti, 2008; Nigg and Holland, 2018). It begins with the recruitment and local accumulation of three key proteins: Polo-like kinase 4 (Plk4), SCL/TAL-1-interrupting locus (STIL), and Spindle assembly abnormal protein 6 (SAS-6). Plk4, is a serine-threonine kinase recruited to the wall of the mother centriole at the onset of biogenesis (Bettencourt-Dias et al., 2005;

Habedanck et al., 2005). This is accomplished through the action of Cep192 and Cep152, together with centriole tethering proteins, which assemble into a cylinder surrounding the entire centriole and into a proximal torus, respectively (Kim et al., 2019; Sonnen et al., 2013; Wei et al., 2020). At first, Plk4 appears to form a discontinuous ring around the mother centriole, which rapidly coalesces into a single focus (Ohta et al., 2014; Park et al., 2019; Takao et al., 2019a; Yamamoto and Kitagawa, 2019). Plk4-dependent phosphorylation of STIL is required for the recruitment of SAS-6 (Kratz et al., 2015; Moyer et al., 2015; Ohta et al., 2018; Park et al., 2020),reviewed in Arquint and Nigg, 2016, which forms the hub of a scaffolding cartwheel structure (Guichard et al., 2017; Kitagawa et al., 2011). The cartwheel is the first identifiable structure by electron microscopy which can be mapped to a budding procentriole (reviewed in Guichard et al., 2018). The concerted action of Plk4, STIL, and SAS-6 thus establishes the site of centriole biogenesis.

Afterwards, the cartwheel directs the nucleation of centriolar microtubules. This highly organised structure consists of ring-like subunits stacked on top of each other (Guichard et al., 2013; Nazarov et al., 2020; van Breugel et al., 2011). Each ring is composed of nine SAS-6 homodimers which interact laterally via their N-terminal head domains (Banterle et al., 2020; Kitagawa et al., 2011; Nievergelt et al., 2018) and project outwards as spokes, giving the cartwheel its distinctive appearance. The distal part of the spokes is probably composed of Cep135/Bld-10, which connects the inner cartwheel hub to the microtubule triplets (Guichard et al., 2017; Guichard et al., 2013; Hiraki et al., 2007)). The cartwheel is an important determinant of centriolar architecture. Indeed, mutations to SAS-6 or to its binding partner Cep135/Bld-10 can generate centrioles with different symmetries or destabilise the centriolar structure, with some slight variations between species (Hiraki et al., 2007; Nakazawa et al., 2007; Noga

et al., 2022; Roque et al., 2012). In humans, the cartwheel is necessary for procentriole formation but its degradation at mitosis indicates it is not necessary for long-term stability of the centriole (Izquierdo et al., 2014; Vorobjev and Chentsov, 1980). Thus, cartwheel formation is a pivotal event in centriole assembly.

Centriole duplication yields two nascent procentrioles positioned orthogonally to the preexisting ones. During S-phase, the microtubule triplets in the newly formed procentrioles elongate slowly via the microtubule-binding protein CPAP (Sharma et al., 2016; Tang et al., 2009). The capping proteins CP110 and Cep97 localise to the distal end of the procentrioles and limit their elongation (Schmidt et al., 2009, reviewed in Sharma et al., 2021). During G2, centriole elongation is maximal and the centrosome recruits further PCM proteins through the action of Polo-like kinase 1 (Plk1), or Polo in the fruit fly, in preparation for mitosis (Alvarez-Rodrigo et al., 2019; Cabral et al., 2019; Lane and Nigg, 1996; Lee and Rhee, 2011; Woodruff et al., 2015). Centriole maturation also involves the acquisition of distal and subdistal appendages, which play a role in tethering the centriole/basal body to the plasma membrane and microtubule anchoring, respectively (Kobayashi and Dynlacht, 2011).

At this point in the cell cycle, both preexisting centrioles are bound by a linker, which is disassembled via an elusive mechanism at the onset of mitosis (reviewed in Agircan et al., 2014). Centrosome separation then occurs as each pair of centrioles, containing the preexisting mother centriole and the fully developed procentriole - or daughter centriole - , migrates to opposite poles of the cell via the action of motor proteins such as Eg5 and dynein (Gaglio et al., 1996; Gönczy et al., 1999; Smith et al., 2011; Tanenbaum et al., 2008). The distance between each migrating centriolar pair peaks during anaphase, at which point the spindle microtubules grow out from the poles and attach to kinetochores; chromosome segregation then ensues. At the end

of mitosis, the mother-daughter pair disengages and the linker is rebuilt (reviewed in Agircan et al., 2014). After cytokinesis, the cycle is restored with each daughter cell inheriting a single pair of centrioles.

The centriole duplication and segregation cycle ensures that centriole number is conserved across cell division in healthy cells. Numeric anomalies in centrioles tend to be deleterious at the cell and organismal levels, and have been associated with several human diseases, such as ciliopathies and cancer (reviewed in Bettencourt-Dias et al., 2011; Godinho et al., 2014). For the purposes of this thesis, I will focus on numeric anomalies resulting in supernumerary centrioles, or centrosome amplification.

Centriole number anomalies and cancer

Ever since the first characterisation of the centrosome and centrioles, the presence of extra centrioles was associated with cancer. Over a century ago, Boveri observed that tumour cells tended to possess extra centrosomes (reviewed in Godinho and Pellman, 2014). These cells could undergo multipolar divisions generating catastrophic aneuploidies, which might trigger tumourigenesis. Nowadays, it is widely accepted there is a strong association between the presence extra centrioles and cancer. First, excess centrioles have been widely reported in cancer tissues and cancer cell lines (Lopes et al., 2018; Marteil et al., 2018,reviewed in Chan, 2011). Second, extra centrioles have been correlated with worse prognosis for a variety of solid and liquid cancers (reviewed Chan, 2011). Third, numeric abnormalities in centrioles have been shown to vary along cancer development in the case of gastroesophageal adenocarcinoma (Lopes et al., 2018). Fourth, extra centrioles have been linked with a variety of cancer-associated phenotypes, such as increased invasiveness and migration in

both cell-autonomous and non-cell-autonomous fashion (Arnandis et al., 2018; Godinho et al., 2014). The question still remains if centriole number abnormalities arise as a consequence of other oncogenic alterations of if they can drive tumourigenesis *per se*. Nevertheless, it is generally accepted that if centrosome amplification can initiate cancer development, it is probably because it promotes chromosomal instability and aneuploidy (reviewed in Godinho and Pellman, 2014).

Aneuploidy, understood as deviations in number of chromosomes to the reference karyotype, is a hallmark of cancer (Weaver and Cleveland, 2008). In contrast with the number of extra centrioles, which is highly heterogeneous within a population of cells and between tumours of the same type, cancer karyotypes tend to feature some recurring aneuploidies. In other words, certain types of tumours tend to display gains or losses of specific chromosomes, despite some variation in the population. Aneuploidy can arise via chromosomal instability, which can be defined as errors leading to the missegregation of entire chromosomes. Such segregation errors can lead to the amplification of oncogenes or to the loss of tumour suppressors, both of which have been commonly observed in cancer (reviewed in Klaasen and Kops, 2022; Nicholson and Cimini, 2015).

The observation that multipolar divisions can induce massive degrees aneuploidy led Boveri to theorise that these erroneous cell divisions could trigger cancer development (reviewed in Godinho and Pellman, 2014). More recent studies indicate that multipolar divisions are extremely lethal (Kwon et al., 2008). These results argue against Boveri's initial hypothesis and it is unlikely that cancer can arise from multipolar divisions. However, chromosome segregation errors in cells carrying extra centrioles do not stem solely from multipolar divisions. In fact, many cells can avoid lethal multipolar divisions by undergoing centrosome clustering (Ganem et al., 2009; Kwon et al., 2008). When

a cell with extra centrioles enters mitosis, a multipolar spindle can arise. This may lead to syntelic microtubule-kinetochore attachments, in which both sister chromatids attach to the same spindle pole. These incorrect attachments can be detected by the spindle-assembly checkpoint, which delays anaphase until these errors are resolved. This allows supernumerary spindle poles to come together through the action of motor proteins, such as dynein and the kinesin KIFC1, thus reinstating spindle bipolarity. However, merotelic microtubule-kinetochore attachments, in which single chromosome is attached to both spindle poles, can also form during the transient multipolar phase. These attachments are not detected by the spindle assembly checkpoint and can result in lagging chromosomes, which are likelier to be incorrectly segregated (Ganem et al., 2009; Kwon et al., 2008). Moreover, the probability of chromosome segregation errors in cells with clustered centrosomes is correlated with the difference in centriole numbers between the two spindle poles. A pole in which there are more centrioles can potentially nucleate more microtubules, and this imbalance favours chromosome segregation errors (Cosenza et al., 2017). Therefore, cells with extra centrioles can generate progeny carrying mild levels of aneuploidy and suppress lethal multipolar divisions via centrosome clustering.

Centrosome clustering has been observed in a variety of cancer cell lines (Moreno-Marin et al., 2023). In addition, inhibition of centrosome clustering can induce massive amounts of cell death, such that it is frequently held as a survival mechanism. There may be some caveats to this notion. First, whereas the offspring of cells with clustered centrosomes tends to be viable, it has been reported that they cannot survive subsequent cell divisions (Cosenza et al., 2017). Thus, centrosome clustering may be insufficient to fully protect cells against centrosome amplification.

Nonetheless, it is still possible that chromosome missegregation

events resulting from centrosome clustering lead to tumour initiation. Cancer can be understood as a process of somatic cell evolution, and as every evolutionary process, it is a game of chance (Ní Leathlobhair and Lenski, 2022). Even if most cells with extra centrioles are eventually purged, there is still some probability that chromosome missegregation leads to the loss of a tumour suppressor or to the amplification of an oncogene. Either of these events can potentially induce cancer development.

Whereas the causal link between extra centrioles and cancer initiation has been repeatedly tested, the subject is still contentious (Coelho et al., 2015; Kulukian et al., 2015; Levine et al., 2017; Serçin et al., 2016; Shoshani et al., 2021; Vitre et al., 2015. Centrosome amplification seems to promote the appearance of certain types of tumours, such as lymphomas and sarcomas, but only in specific contexts (Levine et al., 2017; Shoshani et al., 2021). Although the source of centriole number abnormalities in cancer remains to be characterised, mutations or expression changes in key centriolar genes have been commonly reported. Moreover, certain oncogenes, such as BRCA1 and BRCA2, and other genes that are frequently altered in cancer, such as the master regulator of inflammation NFκB, are known to affect centriole numbers (Ledoux et al., 2013; Starita et al., 2004; Wang et al., 2011). Alterations to such genes could presumably affect centriole overproduction. In addition, mutations in putative "error-correctors" such as p53, which can allow cells with centrosome amplification to survive (Lopes et al., 2018). In order to identify the conditions in which centrosome amplification can lead to tumourigenesis, first it is necessary to understand how extra centrioles are generated and eliminated from the population, and how these two processes interact.

Processes affecting centriole numbers in proliferating cell populations

In this section, I describe the processes affecting centriole numbers in proliferating cell populations. Ensuring fidelity in the centriole duplication and segregation cycle requires that the appearance of numeric anomalies (gain or loss of centrioles) is rare and that any anomalies that arise are "corrected". To do so, the cell depends mainly on tight regulation of centriole biogenesis and negative selection against cells with extra centrioles. I focus on these two processes, with emphasis on how their deregulation can lead to centrosome amplification but briefly discuss other mechanisms that can lead to altered centriole numbers.

Regulation of centriole biogenesis

Although precise estimates of the error rate of centriole duplication are still lacking, it presumably occurs with high fidelity, since cells with atypical centriole numbers are seldom observed *in vivo* (Marshall, 2007). In human cells, centriole number control depends critically on the abundances of Plk4, STIL, and SAS-6 across space and time. Depletion of either protein efficiently inhibits centriole biogenesis and their overexpression tends to induce centriole overproduction (Arquint et al., 2012; Bettencourt-Dias et al., 2005; Kleylein-Sohn et al., 2007; Strnad et al., 2007). Sufficiently high levels of Plk4 can even trigger *de novo* centriole biogenesis, i.e. centriole biogenesis unassisted by preexisting centrioles (Lopes et al., 2015). Therefore, the levels of these proteins must be regulated in order to ensure fidelity in centriole duplication.

Whereas there is some evidence of transcriptional regulation of Plk4 in proliferating cells (Li et al., 2005; Nakamura et al., 2013), these key factors seem to be mostly regulated at the protein level. Several of the identified regulatory mechanisms emerge from the

action of Plk4, STIL, and SAS-6 themselves. For instance, Plk4 can catalyse its own proteasomal degradation by phosphorylating the SCF-Slimb/βTrCP E3 ubiquitin ligase (Cunha-Ferreira et al., 2013; Cunha-Ferreira et al., 2009; Guderian et al., 2010; Rogers et al., n.d.). Indeed, mutations in the Plk4 phosphodegron can prevent its self-destruction, leading to centriole overproduction (Cunha-Ferreira et al., 2013; Guderian et al., 2010). On the other hand, Plk4 can be stabilised through its interaction with STIL (Kratz et al., 2015; Ohta et al., 2018) and, possibly, SAS-6 (Takao et al., 2019a). Plk4 also regulates the abundance of SAS-6 through FBXW5 (Puklowski et al., 2011). In summary, the complex interactions in the Plk4-STIL-SAS-6 network can affect their own abundances.

Differential activity of regulatory mechanisms along the cell cycle allows the cell to restrict centriole biogenesis to S-phase. A recent study found that Plk4, STIL, and SAS-6 levels peak during G1 phase, directly preceding the onset of centriole assembly (Takao et al., 2019a). However, earlier study found that Plk4 reaches it maximum abundance during mitosis (Rogers et al., n.d.). No explanation for this discrepancy has been provided but it may be related with differences in cell types. Furthermore, Plk4 levels at the centrosome undergo similar oscillations during the rapid cell divisions in the *Drosophila* embryo (Aydogan et al., 2018; Nabais et al., 2021). Therefore, the data suggest that the central players of centriole assembly are temporally controlled. At the molecular level, the regulation of Plk4, STIL, and SAS-6 abundance and activity along the cell cyle depends on several components. For example, STIL is degraded by anaphase-promoting complex/cyclosome (APC/C), which inhibits centriole assembly in G2 (Arquint et al., 2012). Furthermore, CDK1 sequesters STIL during late G2 and mitosis, competitively inhibiting its binding to Plk4 (Zitouni et al., 2016).

Taken together, these results indicate the existence of multiple

intrinsic and extrinsic mechanisms that fine-tune the protein abundances of Plk4, STIL, and SAS-6 in space and time. It should be noted that while the cell cycle machinery can restrict centriole assembly, centriole duplication is not indispensable for cell cycle progression. For instance, cells can progress through S-phase even if mother centrioles are laser ablated or if Plk4 is chemically inhibited (Uetake et al., 2007; Wong et al., 2015). Cells completely devoid of centrioles, however, can undergo an irreversible p53-dependent cell cycle arrest (Lambrus et al., 2015; Wong et al., 2015).

The above-mentioned studies convincingly demonstrated the importance of regulating the levels of Plk4, STIL, and SAS-6 for ensuring correct centriole duplication. On a finer scale, however, the issue of centriole number control cannot be reduced to protein homeostasis. The main reason for this lies in the observation that the centrosomal levels of the key regulators of centriole biogenesis are notoriously low (Bauer et al., 2016). Low-abundance proteins tend to be susceptible to noise in gene expression or protein-protein interactions. Therefore, additional mechanisms are required to ensure high fidelity in centriole duplication.

The issues associated with low molecular abundances can be overcome by locally concentrating the key players of centriole biogenesis. For example, concentrating Plk4 at an ectopic location, such as the peroxisome, is sufficient for promoting its self-activation, which is necessary for centriole biogenesis (Lopes et al., 2015). At physiological conditions, the mother centriole has been regarded to act as a concentrator for Plk4 and other components (Nabais et al., 2018). Indeed, when daughter centrioles are laser-ablated or when procentrioles are prematurely disengaged, centrioles are efficiently reduplicated. Contrarily, if the mother centriole is laser-ablated, centrioles are incapable of reduplication (Uetake et al., 2007). In addition to being recruited to the centrosome, centriolar components are concentrated at

a single site around mother centrioles at the onset of centriole biogenesis (Kratz et al., 2015; Ohta et al., 2014; Ohta et al., 2018). As previously mentioned, Plk4 first localises the centrosome as a ring surrounding the mother centriole. Afterwards, the Plk4 ring is resolved to a single dot. The dynamics of this "ring-to-dot" transition are poorly understood but depend on Plk4 kinase activity (Kratz et al., 2015; Ohta et al., 2018; Yamamoto and Kitagawa, 2019).

Two mathematical models were proposed to explain this phenomenon. The first model depicted Plk4 reaction kinetics like a classical Turing reaction-diffusion system: a slowly-diffusing activator (non-phosphorylated Plk4) promotes its own recruitment generates a rapidly-diffusing inhibitor (phosphorylated Plk4), which depletes the activator. Thus, any local increase in non-phosphorylated Plk4 can self-reinforce and generate phosphorylated Plk4, which prevents its accumulation in neighbouring regions (Takao et al., 2019b). The second model achieved a similar result based on two-positive feedback loops: the first, in which accumulation of non-phosphorylated Plk4 can promote the accumulation of labile phosphorylated Plk4, and a second one, in which phosphorylated Plk4 phosphorylates STIL, which in turn stabilises Plk4 (Leda et al., 2018). Both models can robustly reproduce the observation that key centriolar components accumulate in a single focus albeit invoking different mechanisms. Furthermore, the different mathematical approaches complicate comparisons between the two. Encouragingly, a yet unpublished study has attempted to reconcile both models and could potentially shed light on this matter (Wilmott et al., 2023). Notwithstanding, and although the precise molecular mechanism remains undisclosed, it is clear that spatial regulation of centriole biogenesis depends on Plk4 self-activation and self-degradation, as well as Plk4-dependent recruitment of STIL and SAS-6, which in turn, stabilise Plk4.

Control of centriole duplication ultimately requires that this single focus is converted into a single procentriole. The studies above solved this issue by assuming that the space around the preexisting centriole is discretised into separate compartments. Thus, if Plk4, STIL, and/or SAS-6 are stably concentrated into one of these compartments, a single procentriole would arise. It was proposed that these compartments could be established by Cep152 but broad evidence for their evidence is lacking (Takao et al., 2019b). In the absence of these compartments, other mechanisms would be required to ensure that a single procentriole is formed. We addressed this issue previously by proposing a model of cartwheel assembly in which the formation of new cartwheels, and thus, new procentrioles, can be inhibited by elongating the one that formed first. This mechanism could potentially buffer fluctuation in local protein concentrations and allow a single procentriole to arise (Dias Louro et al., 2021).

In conclusion, the regulation of centriole biogenesis depends on the expression levels and intricate reaction dynamics of a small group of proteins. The cell can enact control over spatial and temporal dimensions such that centriole duplication occurs once and only once per cell cycle. The detailed molecular mechanism underlying this process is still poorly understood and existing explanations are not consensual. Notwithstanding, these studies yield some important considerations. First, centriole biogenesis is highly regulated, such that correct duplication should be favoured in most healthy proliferating cells. Second, Plk4, STIL, and SAS-6 determine the number of procentrioles that form in a dose-dependent way. Third, centriole overproduction is a stochastic process leading to a randomly distributed number of newly generated procentrioles depending on the abundance of its building blocks.

While centrosome amplification is typically regarded as an "error" for most proliferating cells, it is part of the developmental

process of certain specialised cell types. For example, terminally differentiated multicilliated cells produce high numbers of centrioles in characteristic structures called deuterosomes (reviewed in Nabais et al., 2018). In addition, mouse olfactory neuron progenitors generate extra centrioles in rosette-like structures commonly observed upon overexpression of Plk4, STIL, and SAS-6. Indeed, these genes were found to be up-regulated during regular brain development, correlating with the increase in centriole numbers (Ching and Stearns, 2020).

It should also be mentioned that there are pathways to centrosome amplification which are not necessarily stochastic nor related to alterations in the levels of key regulators of centriole biogenesis. First, if an otherwise "normal" cell fails to divide, it can restart the cell cycle with four centrioles (reviewed in Normand and King, 2010). Thus, cytokinesis failure can lead to deterministic or quasi-deterministic centriole number increases. Second, DNA damage can lead to a prolonged G2-phase and centrosome amplification (Dodson et al., 2004). Cell cycle arrest following hydroxyurea treatment has also been known to lead to a premature, Plk1-dependent, centriole disengagement, and reduplication in U2OS cells (Lončarek et al., 2010). Therefore, there are other processes aside from random centriole overproduction caused Plk4, STIL, and SAS-6 overexpression that can alter centriole numbers in proliferating cells.

Negative selection against cells with extra centrioles and down-regulation of centriole overproduction

However low the error rate of centriole biogenesis may be, centriole number anomalies can still arise. Maintaining constant centriole numbers requires the existence of so-called "error-correcting mechanisms" to enforce a stable equilibrium state (Marshall, 2007). One such mechanism is negative selection against cells

bearing extra centrioles. As previously mentioned, extra centrioles can induce lethal multipolar cell divisions in the absence of compensatory mechanisms, such as centrosome clustering. Even in the presence of efficient centrosome clustering, as mentioned earlier, cells may not be viable in subsequent cell divisions (Cosenza et al., 2017). Thus, severely aneuploid cell divisions are a form of negative selection acting against cells with extra centrioles.

Whereas the processes listed above arise as a "passive" consequence of centrosome amplification, cells possess mechanisms which can monitor their internal state and respond adequately in the presence of errors. For example, cell cycle checkpoints are intricate pathways that prevent cell cycle progression until certain requirements are met. If detected errors are not repaired, the cell may instead undergo irreversible cell cycle arrest or trigger programmed cell death (Hartwell and Weinert, 1989). For instance, the aforementioned spindle assembly checkpoint delays anaphase until erroneous attachments of chromosomes to spindle poles are resolved (Lara-Gonzalez et al., 2012). In many cases, p53 is responsible for integrating a multitude of stress signals, including DNA damage and genomic instability, and trigger a downstream response (Ozaki and Nakagawara, 2011). The levels of p53 are usually kept in check, most notably, by the Mdm2 ubiquitin ligase (Haupt et al., 1997). When a stress response is triggered, p53 becomes stabilised and activates by translocating to the cell nucleus, where it modulates the expression of target genes, either directly or indirectly (Ozaki and Nakagawara, 2011). Thus, p53 has a near-ubiquitous role in ensuring cellular and organismal homeostasis and it is one of the most commonly mutated genes in cancer. It has been widely reported that p53 becomes active in response to Plk4 overexpression or centrosome amplification both in cell lines and *in vivo*, leading to cell cycle arrest or programmed cell death (Coelho et al., 2015; Holland

et al., 2012; Kulukian et al., 2015; Levine et al., 2017; Serçin et al., 2016; Tkach et al., 2022). Thus, cells with extra centrioles elicit a p53-dependent mechanism which stymies their proliferation.

Recent studies have shed some light on how centriole numeric anomalies lead to p53 activation. In the case of excess centrioles, this response is dependent on the PIDDosome, a protein complex composed of PIDD1, RAIDD1, and its effector protein caspase-2. Following centrosome amplification, PIDD1 is recruited to the centrosome, depending the subdistal appendage protein ODF-2 and the distal appendage protein ANKRD26. This leads to caspase-2 activation, which degrades MDM2. Stabilisation of p53 triggers an irreversible cell cycle arrest (Burigotto et al., 2021; Fava et al., 2017). Loss of extra centrioles on the other hand prevents 53BP1 and USP28 degradation of p53. In turn, this allows p53 to become stabilised and trigger cell cycle arrest via p21 (Fong et al., 2016). Thus, excess centrioles or loss of centrioles can trigger multiple pathways which prevent degradation of p53 and arrest cell proliferation.

In cells harbouring centrosome amplification, it is not known if the strength of negative selection increases with the number of extra centrioles. Several authors have suggested that cells might tolerate low numbers of extra centrioles whereas extremely high numbers may be deleterious (Bloomfield and Cimini, 2023; Levine et al., 2017). It seems plausible that the mechanisms listed above should be more effective in cells carrying extremely high numbers of extra centrioles. For example, a higher number of centrioles could potentially result in a greater number of spindle poles and facilitate multipolar divisions. However, Levine *et al.* reported a significant proliferation defect in cells where Plk4 was mildly overexpressed and most cells contained only a few extra centrioles (Levine et al., 2017). So, it is possible that centrosome amplification induces a fitness cost which is independent of absolute centriole numbers. The data is ultimately insufficient to

support this conclusion.

Aside from negative selection, the cell can theoretically deploy other mechanisms for countering centrosome amplification. For example, negative feedback loops can down-regulate the levels of centriolar proteins and hinder sustained centriole overproduction. An example of this is the self-catalysed instability of Plk4 and its role in regulating the abundances of other centrosomal components (Cunha-Ferreira et al., 2013; Cunha-Ferreira et al., 2009; Puklowski et al., 2011). Several studies also suggest that p53 might repress Plk4 transcription in response to centrosome amplification, but this is yet to be definitively shown (Li et al., 2005; Nakamura et al., 2013). Overall, the data seem to indicate that centrosome amplification and Plk4 overexpression might be ultimately self-inhibitory but experimental backing of this hypothesis is still missing.

In conclusion, there are two main mechanisms that counteract centrosome amplification: negative selection of cells carrying centriole number abnormalities and negative regulation of centriole assembly regulators. Whereas centrioles can be *de facto* eliminated in some contexts, such as during oogenesis (Pimenta-Marques et al., 2016; Łuksza et al., 2013), to the best of our knowledge, this does not seem to occur in typical proliferating cells (Sala et al., 2020). In contrast, several mechanisms can ameliorate the fitness cost of extra centrioles, such as centrosome clustering, as previously discussed. Others such as centrosome inactivation can also prevent multipolar divisions by suppressing microtubule nucleation from additional spindle poles, although their prevalence in cancer is still debatable (Moreno-Marin et al., 2023,reviewed in Sabat-Pośpiech et al., 2019). Finally, asymmetric segregation of extra centrioles can allow one daughter cell to inherit wild-type centriole numbers, allowing it to partially avoid the deleterious effects of centrosome amplification (Baudoin et al., 2020). Therefore, whereas there is pressure in normally cycling

cells to oppose centrosome amplification, several mechanisms can allow cells to tolerate the presence of extra centrioles.

Temporal dynamics of centriole numbers in proliferating cell populations

In healthy human cells, centriole duplication is tightly controlled. In cancerous tissues and derived cell lines, on the other hand, centriole number anomalies are common. Despite these anomalies, cancer cell lines usually display characteristic distributions of centriole numbers per cell (Bettencourt-Dias et al., 2011; Lopes et al., 2018; Marteil et al., 2018), which appear to be maintained by a balance between centriole overproduction and negative selection acting on cells with centrosome amplification, or overproduction-selection balance. This observation suggests centriole numbers are at an equilibrium in proliferating cell populations. In this section I review the literature on centriole number dynamics when this equilibrium is transiently or chronically perturbed, and its consequences for cancer development.

Transient centriole number perturbations

Several lines of evidence indicate that centriole numbers in populations of proliferating cultured cells are recovered upon experimental induction of transient centriole number changes. This observation has led to the proposal that centriole number dynamics are governed by a balance between centriole (over)production and negative selection acting against cells with extra centrioles, which culminates in a population-level equilibrium. In the following I briefly describe studies in which the dynamics of centriole numbers were tracked upon (1) transient Plk4 inhibition (2) transient Plk4 degradation, (3) transient Plk4 overexpression, (4) transient induction of cytokinesis failure.

Transient Plk4 inactivation

Wong *et al.* reported that centriole numbers are kept at equilibrium in a variety of cell lines (Wong et al., 2015). The authors developed centrinone and centrinone B, two drugs that selectively inhibit Plk4 phosphorylation and inhibit centriole duplication. Thus, centrioles were progressively "diluted" as the cells underwent successive rounds of division without centriole duplication. Indeed, when treated with either drug for several generations, centrioles were efficiently depleted from a variety of cell lines. After drug washout, extra centrioles arose *de novo* before the population seemingly resolved back to the initial distribution of centriole numbers per cell. These dynamics were observed in cell lines lacking wild-type p53. Conversely, cell lines containing wild-type p53 underwent irreversible cell-cycle arrest upon centriole loss. This mechanism was found to be independent from DNA damage, chromosome segregation errors, aberrant mitotic duration, or signalling from the Hippo pathway. Thus, centriole number was found to be stable to transient centriole depletion, provided that the cells were capable of proliferation due to loss of wild-type p53 activity.

Transient Plk4 degradation

A caveat of the previous approach was that Plk4 inhibition prevents its own phosphorylation, which targets the protein for proteasomal degradation. Thus, centrinone treatment led to an accumulation of Plk4, which could be responsible for the initial bout of centriole overproduction (Wong et al., 2015). Another study employed a different strategy to deplete cells of centrioles (Lambrus et al., 2015). The authors generated an RPE cell line in which both Plk4 alleles were tagged with an auxin-inducible degron. Upon treatment with indole-3-acetic acid (IAA), Plk4 was rapidly depleted from the centrosome, leading to centriole

duplication failure and an irreversible cell cycle arrest. The observed defect in population growth was partially rescued by p53 knockdown. In cell populations in which p53 expression was down-regulated, Plk4 returned to wild-type levels after IAA washout and, eventually, the distribution of centriole numbers in the population returned to equilibrium. Strikingly, there was an initial excess of centriole overproduction which could scarcely be attributed to higher Plk4 levels. Instead, *de novo centriole* production appeared to be less regulated than canonical duplication, leading to numeric errors, as observed in other studies (Wang et al., 2015).

Transient Plk4 overexpression

To investigate the forces ensuring equilibrium in the distribution of centriole numbers per cell, Sala et al. (Sala et al., 2020) employed an RPE cell line containing a tetracycline-inducible Plk4 overexpression system. After 48h of induction, a significant increase in cells with extra centrioles was observed. The initial distribution of centriole numbers per cell was recovered 12 days after doxycycline washout. The authors proposed four hypotheses that could explain the observation: (1) failure of extra centrioles to undergo centriole duplication, which would lead to their disappearance from the population; (2) asymmetric inheritance of extra centrioles, which could lead to one of the daughter cells inheriting wild-type-like numbers; (3) elimination of extra centrioles; (4) selection against cells containing extra centrioles. Live- and fixed-cell imaging revealed that extra centrioles reliably underwent centriole duplication, that centriole segregation occurred randomly with no particular bias towards asymmetric segregation, and that extra centrioles were stable across cell division, providing evidence against the first three hypotheses. However, the authors observed a strong cell cycle delay/arrest in interphase, which correlated with stabilisation

of nuclear p53, but not a significant increase in cell death. These results suggested that selection was the major force responsible for the observed centriole number dynamics. Notwithstanding, Plk4 levels dropped quickly after doxycycline washout, such that a decrease in the rate of centriole overproduction could also have contributed to the observed behaviour.

Along these lines, another study by Serçin et al. (Serçin et al., 2016) investigated the role of Plk4 overexpression in epidermal tumour development. They generated mice conditionally expressing Plk4-mCherry as previously employed for assessing the consequences of Plk4 overexpression during brain development, which was activated by Cre-induced recombination during embryonic development of the skin epidermis. They observed this system generated cells with extra centrioles *in vitro* after transgene activation. In vivo, Plk4 overexpression led to apoptosis via p53-dependent and independent pathway. Moreover, whereas cells robustly recombined the transgene either during embryogenesis or in surviving adults, the levels of Plk4 overexpression were progressively reduced, suggesting transcriptional silencing of the transgene. Concomitantly, the levels of p53 expression and cell death, as well as centriole numbers, returned to basal levels. Thus, cells carrying an active transgene were counter-selected, but the return to wild-type-like centriole numbers also depended on transcriptional silencing of Plk4, which could be interpreted as a reduction of the overproduction rate. Indeed, extra centrioles disappeared even in p53-knockout mice, suggesting that Plk4 silencing occurred via a different mechanism.

Interestingly, Levine et al. (Levine et al., 2017) transfected mice with a single Plk4 transgene, which leads to a modest Plk4 increase when induced with doxycycline. They observed that when mice were transiently fed with doxycycline for one month, tumours containing cells with extra centrosomes arose. Although

these mice accumulated further errors during tumour development, including some abnormalities indicative of impaired p53 activity, that could have enabled extra centrioles to be maintained in the tumours, these results suggested that there are conditions in which transient Plk4 overexpression can lead to persistent centriole number abnormalities, rather than a return to wild-type numbers. Conversely, a subsequent study employed mice carrying the same transgene and induced transient Plk4 overexpression for two days. This was sufficient to induce a robust increase in the penetrance of cells harbouring extra centrin foci, but these were rapidly lost after doxycycline removal. Taken together, these studies suggested that the level and duration of Plk4 overexpression were critical factors in inducing long-standing centriole number abnormalities.

Transient cytokinesis failure

Finally, I focus on two studies which induced transient centriole number increases using drugs which induce cytokinesis failure. Baudoin et al. (Baudoin et al., 2020) sought to answer how wild-type centriole numbers were recovered after polyploidization. They treated RPE-1 p53-null cells with di-hydrocytochalasin B for 20h, which blocks cytokinesis, thus generating tetraploid cells with extra centrioles. After drug washout, they observed that the initial distribution of centriole numbers was recovered, as previously reported, while tetraploid cells were maintained. The authors reported that both RPE-1 p53-bull and DLD-1 cells that inherited wild-type centriole numbers following an asymmetric division were likelier to survive than the offspring of a symmetric division, where both daughter inherited abnormally high centriole numbers. In contrast, these cells which inherited extra centrioles were prone to undergoing multipolar divisions, cell cycle arrest, or cell death. Although higher sample sizes would be convenient for better estimation of the probability of survival

of cells emerging from asymmetric or symmetric centriole segregation, these results suggest p53-dependent cell cycle arrest is not the sole mechanism responsible for maintaining centriole numbers at a population-level equilibrium.

Similarly, Galofré et al. (Galofré et al., 2020) observed an increase in centrin foci in DLD-1 polyploid cells generated from treatment with di-hydrocytochalasin B and knockdown of ESPL-1 (separase), as well as RKO. The number of centrin foci decreased over four days, regardless of treatment. These dynamics were observed despite a significant fraction of the population maintaining a polyploid state. Therefore, the distribution of centriole numbers per cell in this system was seemingly recovered as observed in other studies.

Taken together, these results support the idea that transient centriole number perturbations are almost invariably resolved towards some preestablished equilibrium. Whereas this phenomenon has been observed for a variety of experimentally-induced perturbations (i.e. Plk4 overexpression, induction of cytokinesis failure, etc.), most of the employed cell lines displayed mostly wild-type-like centriole numbers, initially. Finally, there is data suggesting that negative selection operates to eliminate cells with extra centrioles, through p53-dependent or independent mechanisms. Together with the observation that centriole numbers appear to be characteristic of specific cell lines, these results corroborate the hypothesis that the distribution of centriole numbers per cell is maintained at an equilibrium by overproduction-selection balance.

Chronic centriole number perturbations

Whereas the equilibrium distribution of centriole numbers per cell tends to be reinstated following transient centriole number perturbations, the case with chronic perturbations is more complex. The data on this subject is also more scarce. Below, I

describe several studies where chronic Plk4 overexpression was induced in multiple *in vitro* and *in vivo* systems and its effects on centriole numbers.

Vitre et al. (Vitre et al., 2015) produced mice carrying a Cre-inducible Plk4 overexpression system. When mouse embryonic fibroblasts were infected with adenoviruses harbouring the Cre-recombinase, the authors observed a significant increase in cells with extra centrioles for up to one month. The percentage of these cells in the population peaked five days after infection and plateaued thereafter at 30-40%. Since this system was embryonically lethal, another construct was developed which allowed overexpression of Plk4 when tamoxifen was added to the mice food. After dietary supplementation with tamoxifen for two months, mice exhibited elevated Plk4 mRNA levels but no centriole number abnormalities in the lung or kidney. Skin fibroblasts also showed a reduction in the relative frequency of cells with extra centrioles over 8 days in culture after tamoxifen-induced Cre activation. In contrast, an increase of cells with extra centrioles was observed in the liver. Knockout of one or two p53 alleles allowed cells with extra centrioles to emerge *in vitro*, although a downward trend in their relative frequency was observed over eight days after induction of Plk4 overexpression, and produced cells with extra centrioles two months after tamoxifen supplementation in the kidney. Interestingly, Plk4 expression correlated with the number of knocked-out p53 alelles, indicating that p53 could directly or indirectly affect Plk4 expression, as previously reported. Regardless, neither p53-proficient nor p53-null mice experienced increased tumour formation in the following chronic Plk4 overexpression. In summary, lack of p53 rendered cells more permissive to the occurrence of extra centrioles *in vivo*. However, the trends of reduction in the relative frequency of cells with extra centrioles might indicate p53 loss is insufficient to maintain

centriole number anomalies indefinitely. In addition, was interesting to note that extra centrioles were robustly maintained over time in the first system but not in the second one, suggesting specific behaviours depending on the cell line or overexpression system. However, the matter was not discussed in the paper.

To study the effects of Plk4 overexpression and extra centrioles on epidermal development, Kulukian et al. (Kulukian et al., 2015) developed mice bearing a Cre-inducible construct which enabled Plk4 overexpression by around embryonic day 14.5, which coincides with the onset of skin development. In contrast to Serçin et al. (Serçin et al., 2016), the authors observed that extra centrioles persisted in the skin, but their relative frequency in the population decreased over time. Similarly to Serçin et al. (Serçin et al., 2016), however, the occurrence of extra centrioles was concurrent with an increase in p53 expression levels and higher levels of cell death. The results suggested that extra centrioles could persist in the skin even in the presence of active p53.

In contrast, Coelho *et al.* generated mice carrying a tetracycline-inducible Plk4 overexpression system and reported accelerated formation of sarcomas and lymphomas in the absence of p53. Mice solely overexpressing Plk4 did not suffered from increased tumourigenesis. In both cases, Plk4 overexpression induced centrosome amplification in the pancreas and the epidermis. In contrast to other studies, this did not seem to cause cell death or cell cycle arrest but rather an increase in cell proliferation in pancreatic islets, the suprabasal layer of the epidermis and derived keratinocytes. These effects were potentiated by knocking out p53. Furthermore, the architectural defects in the epidermis were similar to those reported in other studies (Kulukian et al., 2015; Serçin et al., 2016). However, a higher incidence of tumours in Plk4-overexpressing p53-null mice in this study compared to the previous might reflect differences in genetic backgrounds of the mice lines or in the genetic constructs that were used.

In the aforementioned paper by Levine et al. (Levine et al., 2017), Plk4 overexpression was also chronically induced. In mouse embryonic fibroblasts, chronic Plk4 overexpression yielded a significant increase in the relative frequency of cells with extra centrioles over 14 days, albeit decreasing over time. In addition, cell populations chronically overexpressing Plk4 exhibited a proliferation deficit. In mice carrying the Plk4 overexpression system, the authors observed a significant increase in Plk4 expression and centrosome number when doxycycline was added to the food for one and eight months. This effect was observed in the skin, spleen, and thymus. No obvious trend between one and eight-months induction was reported. In a model for intestinal tumour formation, lacking APC, chronic Plk4 overexpression induced a steady increase in the relative frequency of mouse embryonic fibroblasts with extra centrioles, in contrast to the downward trend observed previously. Finally, mice carrying the Plk4 over- expression system chronically fed with doxycycline developed lymphomas and sarcomas where cells containing extra centrioles were widespread. In summary, unlike the paper by Vitre *et al.*, chronic Plk4 overexpression produced persistent occurrence of cells with extra centrioles in various tissues. One of the reported differences between the two studies is the fact that in the study by Vitre et al. (Vitre et al., 2015), induction of the overexpression system yielded extreme levels of Plk4 overexpression and, consequently, extreme centriole number anomalies, whereas in the paper by Levine *et al.*, a lower level of Plk4 overexpression was observed, leading to cells with only a few extra centrioles. Therefore, it is possible that the level of Plk4 overexpression and the rate of centriole overproduction determines the long-term centriole number dynamics. Finally, and in contrast to Serçin *et al.* and Coelho *et al.* (Coelho et al., 2015; Serçin et al., 2016), Plk4 overexpression was sufficient to promote cancer development in mice with proficient p53. Possibly, this is also

due to a lower level of Plk4 overexpression/centrosome amplification. In summary, chronic Plk4 overexpression may or may not lead to chronic centrosome amplification. There is evidence that tissue specificity and the level of Plk4 overexpression might affect these dynamics. Similarly, centrosome amplification *per se* seems to favour formation of sarcomas and lymphomas only in specific contexts. Ultimately, the data are inconclusive and demand further investigation.

In brief, the above studies can allow us to draw the following conclusions: (1) Plk4 overexpression *per se* can induce tumour initiation, namely of sarcomas and lymphomas; (2) induction of tumourigenesis may depend on the level of Plk4 overexpression/penetrance of extra centrioles in the tissues; (3) both chronic and transient Plk4 overexpression (and centriole overproduction) can potentially lead to tumourigenesis; (4) the resulting tumours typically show abnormal karyotypes; (5) tissues show different sensitivities to the presence of extra centrioles; (6) extra centrioles can accelerate the development of skin and intestinal tumours in the presence of additional defects. It is important to note that these results are all pertaining to Plk4 overexpression during early embryogenesis. Investigating other sources of centriole overproduction is crucial to disentangling the effect of overexpressing a particular gene and that of extra centrioles. It would also be important to study the consequences of centriole overproduction during adulthood, which could arguably be a more relevant scenario for assessing tumour formation.

The "population genetics" of centriole numbers

For cells with extra centrioles to play a role in cancer evolution, they must arise, rise in frequency to some extent and remain in a population of cells for some time. Concerning the former, the source of centrosome amplification in tumours or pre-cancerous

lesions remains unknown. Unravelling the molecular basis of centrosome amplification is necessary for understand how it may lead to certain types of tumours rather than others. Concerning their dynamics, however, the studies above indicate that transient centriole number increases and/or low penetrance of centrosome amplification may promote cancer development (Levine et al., 2017; Shoshani et al., 2021). In this regard, evolutionary theory provides a powerful approach for quantifying centriole number dynamics and assessing how the probability of cancer development changes with the relative frequency of cells with centrosome amplification.

Centriole number dynamics are akin to mutation-selection processes

Mutation and selection are regarded as two of the most significant microevolutionary forces in classical evolutionary theory. In population genetics, mutation is a random process which can alter wild-type alleles, thus introducing new variants in a population of individuals. Selection, on the other hand, acts on existing genetic variation, allowing beneficial alleles to increase in frequency while purging deleterious ones. The principle of mutation-selection balance was motivated by observations that in many circumstances mutations are rare and tend to be deleterious. In these conditions, alleles generated by mutation are eliminated by natural selection. When these two forces balance each other out, the relative frequency of mutant alleles in the populations reaches an equilibrium (Hill, 2002).

Mutation-selection processes received extensive mathematical treatment in the advent of population genetics (Fisher, 1930; Haldane, 1937). Of particular interest, the conditions in which mutation-selection balance can hold were derived in simple systems containing few segregating loci. The general intuition is that if mutation rates are too high relative to the fitness cost they

impart, mutations would keep accumulating in the population instead of reaching some stable equilibrium frequency. When this mutation accumulation occurs, the population may experience Muller's ratchet (Muller, 1932), in which the fitness of the population progressively decreases, or mutational meltdown, in which the increased mutational load drives the population to extinction (Lynch et al., 1993). Thus, mutation-selection balance can only be verified under certain conditions - broadly, when selection is strong enough to counter the influx of mutations.

Later, Moran proposed a model of haploid populations featuring an arbitrary number of loci and offered a generalisation of the conditions which allow for mutation-selection balance (Moran, 1977). Independently of Moran, Eigen developed a mathematical model of self-replicating molecules to determine the limits to information transmission at the origin of life (Eigen, 1971). If said molecules were either too large or under sufficiently strong mutational pressure, they would experience too many errors in replication and fail to pass their genetic information to their offspring. The critical value at which genetic information is lost was dubbed the error threshold and it is comparable to the conditions which break mutation-selection balance. In fact, the models proposed by Moran and Eigen were nearly identical, despite stemming from wholly different backgrounds, and allowed for a comprehensive understanding of the dynamics of mutation-selection processes.

Centriole number dynamics bear some similarities to mutation-selection processes. First, the presence of extra centrioles tends to be deleterious. Secondly, centriole overproduction can occur randomly due to stochasticity in gene expression or reaction dynamics. Third, individual cells in a population can potentially harbour variable numbers of centrioles. Therefore, the interaction between centriole overproduction and selection against cells with extra centrioles can be interpreted as a mutation-selection process.

Finally, cell lines tend to have a characteristic distribution of centriole numbers per cell, which is recovered upon transient perturbation. Thus, centriole numbers tend to be kept at an equilibrium, resembling mutation-selection balance.

Based on this analogy with mutation-selection processes, the equilibrium distribution in centriole numbers per cell depends on the centriole overproduction rate and the strength of selection against extra centrioles. Changes in the expression level of centrosomal genes could potentially increase the rate of centriole overproduction, as previously discussed. The establishment of coping mechanisms, such as centrosome clustering, can reduce the fitness cost associated with extra centrioles. Indeed, knocking down or mutating p53 in metaplasia-derived Barrett's oesophagus cell lines improved the viability of cells with centrosome amplification. The relative frequency of cells with extra centrioles rose to the level observed in the succeeding dysplasia stage, in which wild-type p53 activity is compromised (Lopes et al., 2018). Thus, increased rates of centriole overproduction and/or reduced selective pressure against cells with extra centrioles can explain sustained centrosome amplification during cancer development.

In summary, centriole number dynamics are reminiscent of mutation-selection processes. In theory, changes to the equilibrium distributions of centriole numbers per cell in different systems could be explained by changes to the rate of centriole overproduction or to the strength of selection against cells with centrosome amplification. I propose that models of mutation-selection processes can quantitatively explain centriole number variation within and between different populations of cells.

Centrosome amplification is a mutator phenotype

A mutator phenotype is defined as a phenotype which leads to an increased mutation rate. It is typically associated with faulty DNA replication or defects in DNA repair, both of which are

commonly observed in cancer (Loeb et al., 1974). In addition to alterations which increase the mutation rate for point mutations, aneuploidy and chromosomal instability have also been labelled as a mutator phenotype (Holland and Cleveland, 2012). Chromosomal instability refers to certain conditions which potentiate chromosome segregation errors, leading to aneuploidy. Modest levels of chromosome instability allow cells to "explore" the karyotypic space and settle on chromosomal configurations which favour cancer development. Severe chromosomal instability, on the other hand, suppresses tumour formation because it leads to an excessive amount of deleterious copy number changes or because it breaks apart potentially advantageous karyotypes (Nicholson and Cimini, 2015; Weaver and Cleveland, 2008; Weaver et al., 2007).

Since centrosome amplification can induce chromosomal instability it can be considered, by extension, a mutator phenotype. Like chromosomal instability, it has been suggested that low levels of centrosome amplification can favour tumourigenesis whereas high levels hinder tumour formation (Levine et al., 2017) but this hypothesis has not been systematically tested. Moreover, the relationship between centrosome amplification and chromosomal instability is complex and may result in intricate mutator dynamics. First, chromosome segregation errors seem to be favoured when centrioles are unequally distributed by spindle poles, as previously mentioned (Cosenza et al., 2017). Second, even if aneuploidy does arise, the presence of extra centrioles in the daughter cells may inhibit further cell proliferation. As discussed above, this fitness penalty can be seemingly bypassed if the mother cell divides asymmetrically and one daughter cell inherits wild-type centriole numbers (Baudoin et al., 2020). This mechanism allows tetraploid cells to be maintained following cytokinesis failure. Therefore, it is plausible that asymmetric divisions can also allow aneuploid progeny to be maintained.

Taken together, these observations suggest that the number of extra centrioles in a cell and how they are segregated at mitosis might have a crucial role in centrosome amplification-mediated chromosomal instability. To the best of our knowledge, these dynamics were yet to be quantitatively described.

In conclusion, centriole number dynamics share various properties with well-known evolutionary processes. Mutation-selection balance and mutator theories provide a powerful conceptual and quantitative tools for quantifying the forces that affect centriole numbers and their consequences for cell physiology, namely in the generation of aneuploidy. In order to uncover the ultimate implications of centrosome amplification for cancer evolution, it is paramount to understand how the proximal dynamics of centriole numbers in dividing cells. Doing so requires knowledge on how selection acts and on the probability of chromosome segregation errors depend on the number of extra centrioles/level of Plk4 overexpression.

Overview of research work

This thesis revolves around the population dynamics of cells containing abnormally high numbers of centrioles. In the chapter II, we developed mathematical models based on mutation-selection processes and assessed whether or not they can explain quantitatively the observed centriole number distributions in a broad panel of human cancerous and non-cancerous cell lines. We inferred the overproduction rate and strength of selection in our study set and discussed the limitations of parameter estimation based on available data. In chapter III, we extended our previous modelling approach to study how centriole number dynamics depended on centriole segregation and the centrosome cycle. We quantified the probability of asymmetric centriole segregation, in which one daughter cell inherits all extra centrioles, and

explored how the equilibrium centriole number distributions were affected depending on the cell cycle stage at which selection acted. In chapter IV, we devised an experimental evolution setup to answer how the level of centriole overproduction, under chronic induction of Plk4 overexpression, affected the temporal dynamics of centriole number, and other aspects of cell physiology, such as fitness and growth rates. In the final section, we discuss our results in the broader context of centriole number abnormalities in cancer, and the link between cell biology and evolution.

References

Agircan, F. G., Schiebel, E., & Mardin, B. R. (2014). Separate to operate: Control of centrosome positioning and separation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1650), 20130461. https://doi.org/10.1098/rstb.2013.0461

Alvarez-Rodrigo, I., Steinacker, T. L., Saurya, S., Conduit, P. T., Baumbach, J., Novak, Z. A., Aydogan, M. G., Wainman, A., & Raff, J. W. (2019). Evidence that a positive feedback loop drives centrosome maturation in fly embryos (Y. M. Yamashita & A. Akhmanova, Eds.). *eLife*, *8*, e50130. https://doi.org/10.7554/eLife.50130

Arnandis, T., Monteiro, P., Adams, S. D., Bridgeman, V. L., Rajeeve, V., Gadaleta, E., Marzec, J., Chelala, C., Malanchi, I., Cutillas, P. R., & Godinho, S. A. (2018). Oxidative Stress in Cells with Extra Centrosomes Drives Non-Cell-Autonomous Invasion. *Developmental Cell*, *47*(4), 409–424.e9. https://doi.org/10.1016/j.devcel.2018.10.026

Arquint, C., Sonnen, K. F., Stierhof, Y.-D., & Nigg, E. A. (2012). Cell-cycle-regulated expression of STIL controls centriole number in human cells. *Journal of Cell Science*, *125*(5), 1342–1352. https://doi.org/10.1242/jcs.099887

Arquint, C., & Nigg, E. A. (2016). The PLK4–STIL–SAS-6 module at the core of centriole duplication. *Biochemical Society Transactions*, *44*(5), 1253–1263. https://doi.org/10.1042/bst20160116

Aydogan, M. G., Wainman, A., Saurya, S., Steinacker, T. L., Caballe, A., Novak, Z. A., Baumbach, J., Muschalik, N., & Raff, J. W. (2018). A homeostatic clock sets daughter centriole size in flies. *Journal of Cell Biology*, *217*(4), 1233–1248. https://doi.org/10.1083/jcb.201801014

Banterle, N., Nievergelt, A. P., de Buhr, S., Hatzopoulos, G. N., Brillard, C., Andany, S., Hübscher, T., Sorgenfrei, F., Schwarz,

U. S., Gräter, F., Fantner, G. E., & Gönczy, P. (2020). Surface-catalyzed SAS-6 self-assembly directs centriole formation through kinetic and structural mechanisms. *bioRxiv*. https://doi.org/10.1101/2020.09.04.283184

Baudoin, N. C., Nicholson, J. M., Soto, K., Martin, O., Chen, J., & Cimini, D. (2020). Asymmetric clustering of centrosomes defines the early evolution of tetraploid cells. *eLife*, *9*. https://doi.org/10.7554/eLife.54565

Bauer, M., Cubizolles, F., Schmidt, A., & Nigg, E. a. (2016). Quantitative analysis of human centrosome architecture by targeted proteomics and fluorescence imaging. *Embo*, *35*(19), 1–15. https://doi.org/10.15252/embj

Bettencourt-Dias, M., Rodrigues-Martins, A., Carpenter, L., Riparbelli, M., Lehmann, L., Gatt, M. K., Carmo, N., Balloux, F., Callaini, G., & Glover, D. M. (2005). SAK/PLK4 is required for centriole duplication and flagella development. *Current Biology*, *15*(24), 2199–2207. https://doi.org/10.1016/j.cub.2005.11.042

Bettencourt-Dias, M., Hildebrandt, F., Pellman, D., Woods, G., & Godinho, S. A. (2011). Centrosomes and cilia in human disease. *Trends in Genetics*, *27*(8), 307–315. https://doi.org/10.1016/j.tig.2011.05.004

Bloomfield, M., & Cimini, D. (2023). The fate of extra centrosomes in newly formed tetraploid cells: Should i stay, or should i go? *Frontiers in Cell and Developmental Biology*, *11*. https://doi.org/10.3389/fcell.2023.1210983

Boeynaems, S., Alberti, S., Fawzi, N. L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., Tompa, P., & Fuxreiter, M. (2018). Protein phase separation: A new phase in cell biology. *Trends in Cell Biology*, *28*(6), 420–435. https://doi.org/https://doi.org/10.1016/j.tcb.2018.02.004

Breslow, D. K., & Holland, A. J. (2019). Mechanism and Regulation of Centriole and Cilium Biogenesis. *Annual Review*

of Biochemistry, *88*(1), 691–724. https://doi.org/10.1146/annurev-biochem-013118-111153

Brito, D. A., Gouveia, S. M., & Bettencourt-Dias, M. (2012). Deconstructing the centriole: Structure and number control. *Current Opinion in Cell Biology*, *24*(1), 4–13. https://doi.org/10.1016/j.ceb.2012.01.003

Burigotto, M., Mattivi, A., Migliorati, D., Magnani, G., Valentini, C., Roccuzzo, M., Offterdinger, M., Pizzato, M., Schmidt, A., Villunger, A., Maffini, S., & Fava, L. L. (2021). Centriolar distal appendages activate the centrosome-piddosome-p53 signalling axis via ankrd26. *The EMBO Journal*, *40*(4), e104844. https://doi.org/https://doi.org/10.15252/embj.2020104844

Cabral, G., Laos, T., Dumont, J., & Dammermann, A. (2019). Differential requirements for centrioles in mitotic centrosome growth and maintenance. *Developmental Cell*, *50*(3), 355–366.e6. https://doi.org/https://doi.org/10.1016/j.devcel.2019.06.004

Chan, J. Y. (2011). A clinical overview of centrosome amplification in human cancers. *Int. J. Biol. Sci.*, *7*(8), 1122–1144.

Ching, K., & Stearns, T. (2020). Centrioles are amplified in cycling progenitors of olfactory sensory neurons. *PLoS Biology*, *18*(9). https://doi.org/10.1371/journal.pbio.3000852

Coelho, P. A., Bury, L., Shahbazi, M. N., Liakath-Ali, K., Tate, P. H., Wormald, S., Hindley, C. J., Huch, M., Archer, J., Skarnes, W. C., Zernicka-Goetz, M., & Glover, D. M. (2015). Over-expression of Plk4 induces centrosome amplification, loss of primary cilia and associated tissue hyperplasia in the mouse. *Open Biology*, *5*(12). https://doi.org/10.1098/rsob.150209

Cooper, G. M. (2000). *The cell: A molecular approach* (2nd). National Center for Biotechnology InformationÕs Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK9941/

Cosenza, M. R., Cazzola, A., Rossberg, A., Schieber, N. L., Konotop, G., Bausch, E., Slynko, A., Holland-Letz, T., Raab, M. S., Dubash, T., Glimm, H., Poppelreuther, S., Herold-Mende, C., Schwab, Y., & Krämer, A. (2017). Asymmetric centriole numbers at spindle poles cause chromosome missegregation in cancer. *Cell Reports*, *20*(8), 1906–1920. https://doi.org/https://doi.org/10.1016/j.celrep.2017.08.005

Cunha-Ferreira, I., Bento, I., Pimenta-Marques, A., Jana, S. C., Lince-Faria, M., Duarte, P., Borrego-Pinto, J., Gilberto, S., Amado, T., Brito, D., Rodrigues-Martins, A., Debski, J., Dzhindzhev, N., & Bettencourt-Dias, M. (2013). Regulation of autophosphorylation controls PLK4 self-destruction and centriole number. *Current Biology*, *23*(22), 2245–2254. https://doi.org/10.1016/j.cub.2013.09.037

Cunha-Ferreira, I., Rodrigues-Martins, A., Bento, I., Riparbelli, M., Zhang, W., Laue, E., Callaini, G., Glover, D. M., & Bettencourt-Dias, M. (2009). The SCF/Slimb Ubiquitin Ligase Limits Centrosome Amplification through Degradation of SAK/PLK4. *Current Biology*, *19*(1), 43–49. https://doi.org/10.1016/j.cub.2008.11.037

Darwin, C. (1859). *On the origin of species by means of natural selection* [or the Preservation of Favored Races in the Struggle for Life]. Murray.

Dias Louro, M. A., Bettencourt-Dias, M., & Carneiro, J. (2021). A first-takes-all model of centriole copy number control based on cartwheel elongation. *PLOS Computational Biology*, *17*(5), 1–22. https://doi.org/10.1371/journal.pcbi.1008359

Dobzhansky, T. (1973). Nothing in Biology Makes Sense except in the Light of Evolution. *The American Biology Teacher*, *35*(3), 125–129. https://doi.org/10.2307/4444260

Dodson, H., Bourke, E., Jeffers, L. J., Vagnarelli, P., Sonoda, E., Takeda, S., Earnshaw, W. C., Merdes, A., & Morrison, C.

(2004). Centrosome amplification induced by DNA damage occurs during a prolonged G2 phase and involves ATM. *EMBO Journal*, *23*(19), 3864–3873. https://doi.org/10.1038/sj.emboj.7600393

Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Die Naturwissenschaften*, *58*(10), 465–523. https://doi.org/10.1007/BF00623322

Fava, L. L., Schuler, F., Sladky, V., Haschka, M. D., Soratroi, C., Eiterer, L., Demetz, E., Weiss, G., Geley, S., Nigg, E. A., & Villunger, A. (2017). The PIDDosome activates p53 in response to supernumerary centrosomes. *Genes and Development*, *31*(1), 34–45. https://doi.org/10.1101/gad.289728.116

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics (R. A. Fisher, Ed.). *Philosophical Transactions of the Royal Society of London, A*, *222*, 309–368.

Fisher, R. A. (1930). *The genetical theory of natural selection.* Clarendon Press. https://doi.org/10.5962/bhl.title.27468

Fong, C. S., Mazo, G., Das, T., Goodman, J., Kim, M., O'Rourke, B. P., Izquierdo, D., & Tsou, M.-F. B. (2016). 53bp1 and usp28 mediate p53-dependent cell cycle arrest in response to centrosome loss and prolonged mitosis (J. Pines, Ed.). *eLife*, *5*, e16270. https://doi.org/10.7554/eLife.16270

Friedlander, T., Prizak, R., & Barton, N. H. (2017). Evolution of new regulatory functions on biophysically realistic fitness landscapes. *Nature Communications*, *8(1)*, 216. https://doi.org/10.1038/s41467-017-00238-8.

Gaglio, T, Saredi, A, Bingham, J. B., Hasbani, M. J., Gill, S. R., Schroer, T. A., & Compton, D. A. (1996). Opposing motor activities are required for the organization of the mammalian mitotic spindle pole. *Journal of Cell Biology*, *135*(2), 399–414. https://doi.org/10.1083/jcb.135.2.399

Galofré, C., Asensio, E., Ubach, M., Torres, I. M., Quintanilla, I., Castells, A., & Camps, J. (2020). Centrosome reduction

in newly-generated tetraploid cancer cells obtained by separase depletion. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-65975-1

Ganem, N. J., Godinho, S. A., & Pellman, D. (2009). A mechanism linking extra centrosomes to chromosomal instability. *Nature*, *460*(7252), 278–282. https://doi.org/10.1038/nature08136

Godinho, S. A., & Pellman, D. (2014). Causes and consequences of centrosome abnormalities in cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1650). https://doi.org/10.1098/rstb.2013.0467

Godinho, S. A., Picone, R., Burute, M., Dagher, R., Su, Y., Leung, C. T., Polyak, K., Brugge, J. S., Théry, M., & Pellman, D. (2014). Oncogene-like induction of cellular invasion from centrosome amplification. *Nature*, *510*(7503), 167–171. https://doi.org/10.1038/nature13277

Gomes Pereira, S., Dias Louro, M. A., & Bettencourt-Dias, M. (2021). Biophysical and Quantitative Principles of Centrosome Biogenesis and Structure. *Nature Communications*, *37*, 43–63. https://doi.org/10.1038/s41467-017-00238-8.

Gönczy, P. (2015). Centrosomes and cancer: Revisiting a long-standing relationship. *Nature Reviews Cancer*, *15*(11), 639–652. https://doi.org/10.1038/nrc3995

Guderian, G., Westendorf, J., Uldschmid, A., & Nigg, E. A. (2010). Plk4 trans-autophosphorylation regulates centriole number by controlling trcp-mediated degradation. *Journal of Cell Science*, *123*(13), 2163–2169. https://doi.org/10.1242/jcs.068502

Guichard, P., Hamel, V., Le Guennec, M., Banterle, N., Iacovache, I., Nemcíková, V., Flückiger, I., Goldie, K. N., Stahlberg, H., Lévy, D., Zuber, B., & Gönczy, P. (2017). Cell-free reconstitution reveals centriole cartwheel assembly mechanisms. *Nature Communications*, *8*, 1–9. https://doi.org/10.1038/ncomms14813

Guichard, P., Hachet, V., Majubu, N., Neves, A., Demurtas, D., Olieric, N., Fluckiger, I., Yamada, A., Kihara, K., Nishida, Y., Moriya, S., Steinmetz, M. O., Hongoh, Y., & Gönczy, P. (2013). Native architecture of the centriole proximal region reveals features underlying its 9-fold radial symmetry. *Current Biology*, *23*(17), 1620–1628. https://doi.org/10.1016/j.cub.2013.06.061

Guichard, P., Hamel, V., & Gönczy, P. (2018). The Rise of the Cartwheel: Seeding the Centriole Organelle. *BioEssays*, *40*(4). https://doi.org/10.1002/bies.201700241

Gönczy, P., Pichler, S., Kirkham, M., & Hyman, A. A. (1999). Cytoplasmic Dynein Is Required for Distinct Aspects of Mtoc Positioning, Including Centrosome Separation, in the One Cell Stage Caenorhabditis elegans Embryo. *Journal of Cell Biology*, *147*(1), 135–150. https://doi.org/10.1083/jcb.147.1.135

Habedanck, R., Stierhof, Y.-D., Wilkinson, C. J., & Nigg, E. A. (2005). The polo kinase plk4 functions in centriole duplication. *Nature Cell Biology*, *7*(11), 1140–1146. https://doi.org/10.1038/ncb1320

Haldane, J. B. S. (1927). A mathematical theory of natural and artificial selection, Part V: selection and mutation. *Mathematical Proceeding of the Cambridge Philosophical Society*, *23*, 838–844. https://doi.org/10.1017/S0305004100015644

Haldane, J. B. S. (1937). The effect of variation on fitness. *American Naturalist*, *71*, 337–349. https://doi.org/10.1101/623983

Hartwell, L. H., & Weinert, T. A. (1989). Checkpoints: Controls that ensure the order of cell cycle events. *Science*, *246*(4930), 629–634. https://doi.org/10.1126/science.2683079

Haupt, Y., Maya, R., Kazaz, A., & Oren, M. (1997). Mdm2 promotes the rapid degradation of p53. *Nature*, *387*(6630), 296–299. https://doi.org/10.1038/387296a0

Hill, W. (2002). *The Mathematical Theory of Selection, Recombination and Mutation. R. Burger* (Vol. 79). Wiley.

Hiraki, M., Nakazawa, Y., Kamiya, R., & Hirono, M. (2007). Bld10p constitutes the cartwheel-spoke tip and stabilizes the 9-fold symmetry of the centriole. *Current Biology*, *17*(20), 1778–1783. https://doi.org/https://doi.org/10.1016/j.cub.2007.09.021

Holland, A. J., & Cleveland, D. W. (2012). Losing balance: The origin and impact of aneuploidy in cancer. *EMBO Rep.*, *13*(6), 501–514.

Holland, A. J., Fachinetti, D., Zhu, Q., Bauer, M., Verma, I. M., Nigg, E. A., & Cleveland, D. W. (2012). The autoregulated instability of polo-like kinase 4 limits centrosome duplication to once per cell cycle. *Genes Dev.*, *26*(24), 2684–2689.

Huxley, J., Pigliucci, M., & Muller, G. (2010). *Evolution, the definitive edition: The modern synthesis*. MIT Press. https://books.google.pt/books?id=FxtFAQAAIAAJ

Hyman, A. A., Weber, C. A., & Jülicher, F. (2014). Liquid-liquid phase separation in biology [PMID: 25288112]. *Annual Review of Cell and Developmental Biology*, *30*(1), 39–58. https://doi.org/10.1146/annurev-cellbio-100913-013325

Islas-Morales, P. F., Jime Nez-Garcia, L. F., Mosqueira-Santillan, M., & Voolstra, C. R. (2021). Evolutionary cell biology (ECB): Lessons, challenges, and opportunities for the integrative study of cell evolution. *J. Biosci.*, *46*(1).

Izquierdo, D., Wang, W.-J., Uryu, K., & Tsou, M.-F. (2014). Stabilization of cartwheel-less centrioles for duplication requires cep295-mediated centriole-to-centrosome conversion. *Cell Reports*, *8*(4), 957–965. https://doi.org/https://doi.org/10.1016/j.celrep.2014.07.022

Jana, S. C. (2020). Centrosome structure and biogenesis: Variations on a theme? *Semin Cell Dev Biol.*, *110*, 123–138. https://doi.org/10.1016/j.semcdb.2020.10.014

Karsenti, E. (2008). Self-organization in cell biology: A brief history. https://doi.org/10.1038/nrm2357

Kim, T.-S., Zhang, L., Il Ahn, J., Meng, L., Chen, Y., Lee, E., Bang, J. K., Lim, J. M., Ghirlando, R., Fan, L., Wang, Y.-X., Kim, B. Y., Park, J.-E., & Lee, K. S. (2019). Molecular architecture of a cylindrical self-assembly at human centrosomes. *Nature Communications*, *10*(1), 1151. https://doi.org/10.1038/s41467-019-08838-2

Kitagawa, D., Vakonakis, I., Olieric, N., Hilbert, M., Keller, D., Olieric, V., Bortfeld, M., Erat, M. C., Flückiger, I., Gönczy, P., & Steinmetz, M. O. (2011). Structural basis of the 9-fold symmetry of centrioles. *Cell*, *144*(3), 364–375. https://doi.org/10.1016/j.cell.2011.01.008

Klaasen, S. J., & Kops, G. J. P. L. (2022). Chromosome inequality: Causes and consequences of non-random segregation errors in mitosis and meiosis. *Cells*, *11*(22). https://doi.org/10.3390/cells11223564

Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Habedanck, R., Stierhof, Y.-D., & Nigg, E. A. (2007). Plk4-induced centriole biogenesis in human cells. *Developmental Cell*, *13*(2), 190–202. https://doi.org/https://doi.org/10.1016/j.devcel.2007.07.002

Kobayashi, T., & Dynlacht, B. D. (2011). Regulating the transition from centriole to basal body. *J. Cell Biol.*, *193*(3), 435–444.

Kratz, A. S., Bärenz, F., Richter, K. T., & Hoffmann, I. (2015). Plk4-dependent phosphorylation of STIL is required for centriole duplication. *Biology Open*, *4*(3), 370–377. https://doi.org/10.1242/bio.201411023

Kulukian, A., Holland, A. J., Vitrec, B., Naika, S., Cleveland, D. W., & Fuchsa, E. (2015). Epidermal development, growth control, and homeostasis in the face of centrosome amplification. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(46), E6311–E6320. https://doi.org/10.1073/pnas.1518376112

Kwon, M., Godinho, S. A., Chandhok, N. S., Ganem, N. J., Azioune, A., Thery, M., & Pellman, D. (2008). Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes. *Genes and Development*, *22*(16), 2189–2203. https://doi.org/10.1101/gad.1700908

Lambrus, B. G., Uetake, Y., Clutario, K. M., Daggubati, V., Snyder, M., Sluder, G., & Holland, A. J. (2015). P53 protects against genome instability following centriole duplication failure. *Journal of Cell Biology*, *210*(1), 63–77. https://doi.org/10.1083/jcb.201502089

Lane, H. A., & Nigg, E. A. (1996). Antibody microinjection reveals an essential role for human polo-like kinase 1 (Plk1) in the functional maturation of mitotic centrosomes. *Journal of Cell Biology*, *135*(6), 1701–1713. https://doi.org/10.1083/jcb.135.6.1701

Lara-Gonzalez, P., Westhorpe, F. G., & Taylor, S. S. (2012). The spindle assembly checkpoint. *Current Biology*, *22*(22), R966–R980. https://doi.org/https://doi.org/10.1016/j.cub.2012.10.006

Leda, M., Holland, A. J., & Goryachev, A. B. (2018). Autoamplification and Competition Drive Symmetry Breaking: Initiation of Centriole Duplication by the PLK4-STIL Network. *iScience*, *8*, 222–235. https://doi.org/10.1016/j.isci.2018.10.003

Ledoux, A. C., Sellier, H., Gillies, K., Iannetti, A., James, J., & Perkins, N. D. (2013). NFκB regulates expression of polo-like kinase 4. *Cell Cycle*, *12*(18), 3052–3062.

Lee, K., & Rhee, K. (2011). PLK1 phosphorylation of pericentrin initiates centrosome maturation at the onset of mitosis. *Journal of Cell Biology*, *195*(7), 1093–1101. https://doi.org/10.1083/jcb.201106093

Levine, M. S., Bakker, B., Boeckx, B., Moyett, J., Lu, J., Vitre, B., Spierings, D. C., Lansdorp, P. M., Cleveland, D. W., Lambrechts, D., Foijer, F., & Holland, A. J. (2017). Centrosome

Amplification Is Sufficient to Promote Spontaneous Tumorigenesis in Mammals. *Developmental Cell*, *40*(3), 313–322.e5. https://doi.org/10.1016/j.devcel.2016.12.022

Li, J., Tan, M., Li, L., Pamarthy, D., Lawrence, T. S., & Sun, Y. (2005). SAK, a new polo-like kinase, is transcriptionally repressed by p53 and induces apoptosis upon RNAi silencing. *Neoplasia*, *7*(4), 312–323.

Loeb, L. A., Springgate, C. F., & Battula, N. (1974). Errors in DNA Replication as a Basis of Malignant Changes1. *Cancer Research*, *34*(9), 2311–2321.

Lončarek, J., Hergert, P., & Khodjakov, A. (2010). Centriole reduplication during prolonged interphase requires procentriole maturation governed by plk1. *Current Biology*, *20*(14), 1277–1282. https://doi.org/10.1016/j.cub.2010.05.050

Lopes, C. A. M., Jana, S. C., Cunha-Ferreira, I., Zitouni, S., Bento, I., Duarte, P., Gilberto, S., Freixo, F., Guerrero, A., Francia, M., Lince-Faria, M., Carneiro, J., & Bettencourt-Dias, M. (2015). PLK4 trans-Autoactivation Controls Centriole Biogenesis in Space. *Developmental Cell*, *35*(2), 222–235. https://doi.org/10.1016/j.devcel.2015.09.020

Lopes, C. A., Mesquita, M., Cunha, A. I., Cardoso, J., Carapeta, S., Laranjeira, C., Pinto, A. E., Pereira-Leal, J. B., Dias-Pereira, A., Bettencourt-Dias, M., & Chaves, P. (2018). Centrosome amplification arises before neoplasia and increases upon p53 loss in tumorigenesis. *Journal of Cell Biology*, *217*(7), 2353–2363. https://doi.org/10.1083/jcb.201711191

Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. *PNAS*, *104*, 8597–604. https://doi.org/10.1073/pnas.0702207104

Lynch, M., Bürger, R., Butcher, D., & Gabriel, W. (1993). The Mutational Meltdown in Asexual Populations. *Journal of Heredity*, *84*(5), 339–344. https://doi.org/10.1093/oxfordjournals.jhered.a111354

Lynch, M., Field, M. C., Goodson, H. V., Malik, H. S., Pereira-Leal, J. B., Roos, D. S., P., T. A., & Sazer, S. (2014). A Theoretical Framework for Evolutionary Cell Biology. *Evolutionary cell biology: Two origins, one objective*, *111(48)*, 16990–16994. https://doi.org/10.1073/pnas.1415861111

Lynch, M., Schavemaker, P. E., L., T. J., Y., Hao, & A., P. (2022). Evolutionary bioenergetics of cilliates. *Journal of Eukaryotic Microbiology*, *69(5):e12934*. https://doi.org/https://doi.org/10.1111/jeu.12934

Marshall, W. F. (2007). Stability and robustness of an organelle number control system: modeling and measuring homeostatic regulation of centriole abundance. *Biophysical journal*, *93*(5), 1818–1833. https://doi.org/10.1529/biophysj.107.107052

Marteil, G., Guerrero, A., Vieira, A. F., De Almeida, B. P., Machado, P., Mendonça, S., Mesquita, M., Villarreal, B., Fonseca, I., Francia, M. E., Dores, K., Martins, N. P., Jana, S. C., Tranfield, E. M., Barbosa-Morais, N. L., Paredes, J., Pellman, D., Godinho, S. A., & Bettencourt-Dias, M. (2018). Overelongation of centrioles in cancer promotes centriole amplification and chromosome missegregation. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-03641-x

Mazzarello, P. (1999). A unifying concept: The history of cell theory. *Nature Cell Biology*, *1*(1), E13–E15. https://doi.org/10.1038/8964

Moran, P. A. (1977). Global stability of genetic systems governed by mutation and selection. II. *Mathematical Proceedings of the Cambridge Philosophical Society*, *81*(3), 435–441. https://doi.org/10.1017/S0305004100053500

Moreno-Marin, N., Marteil, G., Fresmann, N. C., de Almeida, B., Dores, K., Fragoso, R., Cardoso, J., Pereira-Leal, J., Barata, J., Godinho, S., Barbosa-Morais, N., & Bettencourt-Dias,

M. (2023). High prevalence and dependence of centrosome clustering in mesenchymal tumors and leukemia. *bioRxiv*. https://doi.org/10.1101/2023.03.13.532472

Moyer, T. C., Clutario, K. M., Lambrus, B. G., Daggubati, V., & Holland, A. J. (2015). Binding of STIL to Plk4 activates kinase activity to promote centriole assembly. *Journal of Cell Biology*, *209*(6), 863–878. https://doi.org/10.1083/jcb.201502088

Muller, H. J. (1932). Some genetic aspects of sex. *The American Naturalist*, *66*(703), 118–138. https://doi.org/10.1086/280418

Nabais, C., Gomes Pereira, S., & Bettencourt-Dias, M. (2018). Noncanonical Biogenesis of Centrioles and Basal Bodies. *Cold Spring Harb Symp Quant Biol.*, *82*, 123–135. https://doi.org/10.1101/sqb.2017.82.034694

Nabais, C., Peneda, C., & Bettencourt-Dias, M. (2020). Evolution of centriole assembly. *Current Biology*, *30(10)*, R494–R502. https://doi.org/10.1016/j.cub.2020.02.036

Nabais, C., Pessoa, D., de Carvalho, J., van Zanten, T., Duarte, P., Mayor, S., Carneiro, J., Telley, I. A., & Bettencourt-Dias, M. (2021). Plk4 triggers autonomous de novo centriole biogenesis and maturation. *Journal of Cell Biology*, *220*(5), e202008090. https://doi.org/10.1083/jcb.202008090

Nakamura, T., Saito, H., & Takekawa, M. (2013). Sapk pathways and p53 cooperatively regulate plk4 activity and centrosome integrity under stress. *Nature Communications*, *4*(1), 1775. https://doi.org/10.1038/ncomms2752

Nakazawa, Y., Hiraki, M., Kamiya, R., & Hirono, M. (2007). SAS-6 is a Cartwheel Protein that Establishes the 9-Fold Symmetry of the Centriole. *Current Biology*, *17*(24), 2169–2174. https://doi.org/10.1016/j.cub.2007.11.046

Nazarov, S., Bezler, A., Hatzopoulos, G. N., Villímová, V. N., Demurtas, D., Guennec, M. L., Guichard, P., & Gönczy, P. (2020). Novel features of centriole polarity and cartwheel

stacking revealed by cryo-tomography. *bioRxiv*, 2020.07.17.208082. https://doi.org/10.1101/2020.07.17.208082

Ní Leathlobhair, M., & Lenski, R. E. (2022). Population genetics of clonally transmissible cancers. *Nature Ecology & Evolution*, *6*(8), 1077–1089. https://doi.org/10.1038/s41559-022-01790-3

Nicholson, J. M., & Cimini, D. (2015). Link between aneuploidy and chromosome instability. *International Review of Cell and Molecular Biology*, *315*, 299–317. https://doi.org/10.1016/bs.ircmb.2014.11.002

Nievergelt, A. P., Banterle, N., Andany, S. H., Gönczy, P., & Fantner, G. E. (2018). High-speed photothermal off-resonance atomic force microscopy reveals assembly routes of centriolar scaffold protein SAS-6. *Nature Nanotechnology*, *13*(8), 696–701. https://doi.org/10.1038/s41565-018-0149-4

Nigg, E. A., & Holland, A. J. (2018). Once and only once: Mechanisms of centriole duplication and their deregulation in diseases. *Nature Reviews Molecular Cell Biology*, *19*(5), 297–312. https://doi.org/10.1038/nrm.2017.127

Noga, A., Horii, M., Goto, Y., Toyooka, K., Ishikawa, T., & Hirono, M. (2022). Bld10p/cep135 determines the number of triplets in the centriole independently of the cartwheel. *The EMBO Journal*, *41*(20), e104582. https://doi.org/https://doi.org/10.15252/embj.2020104582

Normand, G., & King, R. W. (2010). Understanding cytokinesis failure. *Adv. Exp. Med. Biol.*, *676*, 27–55.

Ohta, M., Ashikawa, T., Nozaki, Y., Kozuka-Hata, H., Goto, H., Inagaki, M., Oyama, M., & Kitagawa, D. (2014). Direct interaction of Plk4 with STIL ensures formation of a single procentriole per parental centriole. *Nature communications*, *5*, 5267. https://doi.org/10.1038/ncomms6267

Ohta, M., Watanabe, K., Ashikawa, T., Nozaki, Y., Yoshiba, S., Kimura, A., & Kitagawa, D. (2018). Bimodal Binding of STIL to Plk4 Controls Proper Centriole Copy Number.

Cell Reports, *23*(11), 3160–3169.e4. https://doi.org/10.1016/j.celrep.2018.05.030

Ozaki, T., & Nakagawara, A. (2011). Role of p53 in cell death and human cancers. *Cancers*, *3*(1), 994–1013. https://doi.org/10.3390/cancers3010994

Park, J.-E., Meng, L., Ryu, E. K., Nagashima, K., Baxa, U., Bang, J. K., & Lee, K. S. (2020). Autophosphorylation-induced self-assembly and stil-dependent reinforcement underlie plk4's ring-to-dot localization conversion around a human centriole [PMID: 33323015]. *Cell Cycle*, *19*(24), 3419–3436. https://doi.org/10.1080/15384101.2020.1843772

Park, J.-E., Zhang, L., Bang, J. K., Andresson, T., DiMaio, F., & Lee, K. S. (2019). Phase separation of polo-like kinase 4 by autoactivation and clustering drives centriole biogenesis. *Nature Communications*, *10*(1), 4959. https://doi.org/10.1038/s41467-019-12619-2

Pimenta-Marques, A., Bento, I., Lopes, C. A. M., Duarte, P., Jana, S. C., & Bettencourt-Dias, M. (2016). A mechanism for the elimination of the female gamete centrosome in drosophila melanogaster. *Science*, *353*(6294), aaf4866. https://doi.org/10.1126/science.aaf4866

Puklowski, A., Homsi, Y., Keller, D., May, M., Chauhan, S., Kossatz, U., Grünwald, V., Kubicka, S., Pich, A., Manns, M. P., Hoffmann, I., Gönczy, P., & Malek, N. P. (2011). The SCF-FBXW5 E3-ubiquitin ligase is regulated by PLK4 and targets HsSAS-6 to control centrosome duplication. *Nature Cell Biology*, *13*(8), 1004–1009. https://doi.org/10.1038/ncb2282

Rogers, G. C., Rusan, N. M., Roberts, D. M., Peifer, M., & Rogers, S. L. (n.d.). The scf slimb ubiquitin ligase regulates plk4/sak levels to block centriole reduplication. *184*(2), 225–239. https://doi.org/10.1083/jcb.200808049

Roque, H., Wainman, A., Richens, J., Kozyrska, K., Franz, A., & Raff, J. W. (2012). Drosophila Cep135/Bld10 maintains

proper centriole structure but is dispensable for cartwheel formation. *Journal of Cell Science*, *125*(23), 5881–5886. https://doi.org/10.1242/jcs.113506

Sabat-Pośpiech, D., Fabian-Kolpanowicz, K., Prior, I. A., Coulson, J. M., & Fielding, A. B. (2019). Targeting centrosome amplification, an Achilles' heel of cancer. *Biochemical Society Transactions*, *47*(5), 1209–1222. https://doi.org/10.1042/BST20190034

Sala, R., Farrell, K., & Stearns, T. (2020). Growth disadvantage associated with centrosome amplification drives population-level centriole number homeostasis. *Molecular Biology of the Cell*, mbc.E19–04–0195. https://doi.org/10.1091/mbc.e19-04-0195

Schmidt, T. I., Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Lavoie, S. B., Stierhof, Y.-D., & Nigg, E. A. (2009). Control of centriole length by cpap and cp110. *Current Biology*, *19*(12), 1005–1011. https://doi.org/https://doi.org/10.1016/j.cub.2009.05.016

Serçin, Ö., Larsimont, J. C., Karambelas, A. E., Marthiens, V., Moers, V., Boeckx, B., Le Mercier, M., Lambrechts, D., Basto, R., & Blanpain, C. (2016). Transient PLK4 overexpression accelerates tumorigenesis in p53-deficient epidermis. *Nature Cell Biology*, *18*(1), 100–110. https://doi.org/10.1038/ncb3270

Sharma, A., Aher, A., Dynes, N., Frey, D., Katrukha, E., Jaussi, R., Grigoriev, I., Croisier, M., Kammerer, R., Akhmanova, A., Gönczy, P., & Steinmetz, M. (2016). Centriolar cpap/sas-4 imparts slow processive microtubule growth. *Developmental Cell*, *37*(4), 362–376. https://doi.org/https://doi.org/10.1016/j.devcel.2016.04.024

Sharma, A., Olieric, N., & Steinmetz, M. O. (2021). Centriole length control [Centrosomal Organization and Assemblies Folding and Binding]. *Current Opinion in Structural*

Biology, *66*, 89–95. https://doi.org/https://doi.org/10.1016/j.sbi.2020.10.011

Shoshani, O., Bakker, B., de Haan, L., Tijhuis, A. E., Wang, Y., Kim, D. H., Maldonado, M., Demarest, M. A., Artates, J., Zhengyu, O., Mark, A., Wardenaar, R., Sasik, R., Spierings, D. C. J., Vitre, B., Fisch, K., Foijer, F., & Cleveland, D. W. (2021). Transient genomic instability drives tumorigenesis through accelerated clonal evolution. *Genes Dev.*, *35*(15-16), 1093–1108.

Smith, E., Hégarat, N., Vesely, C., Roseboom, I., Larch, C., Streicher, H., Straatman, K., Flynn, H., Skehel, M., Hirota, T., Kuriyama, R., & Hochegger, H. (2011). Differential control of eg5-dependent centrosome separation by plk1 and cdk1. *The EMBO Journal*, *30*(11), 2233–2245. https://doi.org/https://doi.org/10.1038/emboj.2011.120

Smocovitis, V. B. (1992). Unifying biology: The evolutionary synthesis and evolutionary biology. *Journal of the History of Biology*, *25*(1), 1–65. Retrieved September 9, 2023, from http://www.jstor.org/stable/4331201

Sonnen, K. F., Gabryjonczyk, A.-M., Anselm, E., Stierhof, Y.-D., & Nigg, E. A. (2013). Human Cep192 and Cep152 cooperate in Plk4 recruitment and centriole duplication. *Journal of Cell Science*, *126*(14), 3223–3233. https://doi.org/10.1242/jcs.129502

Starita, L. M., Machida, Y., Sankaran, S., Elias, J. E., Griffin, K., Schlegel, B. P., Gygi, S. P., & Parvin, J. D. (2004). BRCA1-dependent ubiquitination of gamma-tubulin regulates centrosome number. *Mol. Cell. Biol.*, *24*(19), 8457–8466.

Strnad, P., Leidel, S., Vinogradova, T., Euteneuer, U., Khodjakov, A., & Gönczy, P. (2007). Regulated HsSAS-6 Levels Ensure Formation of a Single Procentriole per Centriole during the Centrosome Duplication Cycle. *Developmental Cell*, *13*(2), 203–213. https://doi.org/10.1016/j.devcel.2007.07.004

Takao, D., Watanabe, K., Kuroki, K., & Kitagawa, D. (2019a). Feedback loops in the Plk4–STIL–HsSAS6 network coordinate site selection for procentriole formation. *Biology Open*, *8*(9). https://doi.org/10.1242/bio.047175

Takao, D., Yamamoto, S., & Kitagawa, D. (2019b). A theory of centriole duplication based on self-organized spatial pattern formation. *Journal of Cell Biology*, *218*(11), 3537–3547. https://doi.org/10.1083/jcb.201904156

Tanenbaum, M. E., Macůrek, L., Galjart, N., & Medema, R. H. (2008). Dynein, lis1 and clip-170 counteract eg5-dependent centrosome separation during bipolar spindle assembly. *The EMBO Journal*, *27*(24), 3235–3245. https://doi.org/https://doi.org/10.1038/emboj.2008.242

Tang, C.-J. C., Fu, R.-H., Wu, K.-S., Hsu, W.-B., & Tang, T. K. (2009). Cpap is a cell-cycle regulated protein that controls centriole length. *Nature Cell Biology*, *11*(7), 825–831. https://doi.org/10.1038/ncb1889

Tkach, J. M., Philip, R., Sharma, A., Strecker, J., Durocher, D., & Pelletier, L. (2022). Global cellular response to chemical perturbation of plk4 activity and abnormal centrosome number (J. Lüders & A. Akhmanova, Eds.). *eLife*, *11*, e73944. https://doi.org/10.7554/eLife.73944

Uetake, Y., Loncarek, J., Nordberg, J. J., English, C. N., La Terra, S., Khodjakov, A., & Sluder, G. (2007). Cell cycle progression and de novo centriole assembly after centrosomal removal in untransformed human cells. *Journal of Cell Biology*, *176*(2), 173–182. https://doi.org/10.1083/jcb.200607073

van Breugel, M, Hirono, M, Andreeva, A, Yanagisawa, H. A., Yamaguchi, S, Nakazawa, Y, Morgner, N, Petrovich, M, Ebong, I. O., Robinson, C. V., Johnson, C. M., Veprintsev, D, & Zuber, B. (2011). Structures of SAS-6 suggest its organization in centrioles. *Science*, *331*(6021), 1196–1199. https://doi.org/10.1126/science.1199325

Vitre, B., Holland, A. J., Kulukian, A., Shoshani, O., Hirai, M., Wanga, Y., Maldonado, M., Cho, T., Boubaker, J., Swing, D. A., Tessarollo, L., Evans, S. M., Fuchs, E., & Cleveland, D. W. (2015). Chronic centrosome amplification without tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(46), E6321–E6330. https://doi.org/10.1073/pnas.1519388112

Vorobjev, I. A., & Chentsov, Y. S. (1980). The ultrastructure of centriole in mammalian tissue culture cells. *Cell Biol. Int. Rep.*, *4*(11), 1037–1044.

Wang, H.-F., Takenaka, K., Nakanishi, A., & Miki, Y. (2011). BRCA2 and nucleophosmin coregulate centrosome amplification and form a complex with the rho effector kinase ROCK2. *Cancer Res.*, *71*(1), 68–77.

Wang, W. J., Acehan, D., Kao, C. H., Jane, W. N., Uryu, K., & Tsou, M. F. B. (2015). De novo centriole formation in human cells is error-prone and does not require SAS-6 self-assembly. *eLife*, *4*(NOVEMBER2015), 1–13. https://doi.org/10.7554/eLife.10586

Weaver, B. A., & Cleveland, D. W. (2008). The aneuploidy paradox in cell growth and tumorigenesis. *Cancer Cell*, *14*(6), 431–433. https://doi.org/https://doi.org/10.1016/j.ccr.2008.11.011

Weaver, B. A., Silk, A. D., Montagna, C., Verdier-Pinard, P., & Cleveland, D. W. (2007). Aneuploidy acts both oncogenically and as a tumor suppressor. *Cancer Cell*, *11*(1), 25–36. https://doi.org/https://doi.org/10.1016/j.ccr.2006.12.003

Wei, Z., Kim, T.-S., Ahn, J. I., Meng, L., Chen, Y., Ryu, E. K., Ku, B., Zhou, M., Kim, S. J., Bang, J. K., van Deursen, J. M., Park, J.-E., & Lee, K. S. (2020). Requirement of the cep57-cep63 interaction for proper cep152 recruitment and centriole duplication [PMID: 32152252]. *Molecular*

and Cellular Biology, *40*(10), e00535–19. https://doi.org/10.1128/MCB.00535-19

Welch, J. J. (2017). What's wrong with evolutionary biology? *Biology & Philosophy*, *32*(2), 263–279. https://doi.org/10.1007/s10539-016-9557-8

Wilmott, Z. M., Goriely, A., & Raff, J. W. (2023). A simple turing reaction-diffusion model can explain how mother centrioles break symmetry to generate a single daughter. *bioRxiv*. https://doi.org/10.1101/2023.02.02.526828

Wong, Y. L., Anzola, J. V., Davis, R. L., Yoon, M., Motamedi, A., Kroll, A., Seo, C. P., Hsia, J. E., Kim, S. K., Mitchell, J. W., Mitchell, B. J., Desai, A., Gahman, T. C., Shiau, A. K., & Oegema, K. (2015). Reversible centriole depletion with an inhibitor of Polo-like kinase 4. *Science*, *348*(6239), 1155–1160. https://doi.org/10.1126/science.aaa5111

Woodruff, J. B., Wueseke, O., Viscardi, V., Mahamid, J., Ochoa, S. D., Bunkenborg, J., Widlund, P. O., Pozniakovsky, A., Zanin, E., Bahmanyar, S., Zinke, A., Hong, S. H., Decker, M., Baumeister, W., Andersen, J. S., Oegema, K., & Hyman, A. A. (2015). Regulated assembly of a supramolecular centrosome scaffold in vitro. *Science*, *348*(6236), 808–812. https://doi.org/10.1126/science.aaa3923

Wright, S. (1938). Size of population and breeding structure in relation to evolution. *Science*, *87*, 337–349.

Yamamoto, S., & Kitagawa, D. (2019). Self-organization of Plk4 regulates symmetry breaking in centriole duplication. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-09847-x

Zitouni, S., Francia, M., Leal, F., Montenegro Gouveia, S., Nabais, C., Duarte, P., Gilberto, S., Brito, D., Moyer, T., Kandels-Lewis, S., Ohta, M., Kitagawa, D., Holland, A., Karsenti, E., Lorca, T., Lince-Faria, M., & Bettencourt-Dias, M. (2016). Cdk1 prevents unscheduled plk4-stil complex assembly in centriole biogenesis. *Current Biology*, *26*(9), 1127–1137.

https://doi.org/https://doi.org/10.1016/j.cub.2016.03.055

Łuksza, M., Queguigner, I., Verlhac, M.-H., & Brunet, S. (2013). Rebuilding mtocs upon centriole loss during mouse oogenesis. *Developmental Biology*, *382*(1), 48–56. https://doi.org/https://doi.org/10.1016/j.ydbio.2013.07.029

Chapter 2

Patterns of selection against centrosome amplification

in human cell lines

Marco António Dias Louro[1], Mónica Bettencourt-Dias[1], Claudia Bank[1,2]

1. Instituto Gulbenkian de Ciência, Oeiras, Portugal

2. Institute of Ecology and Evolution, University of Bern, Bern, Switzerland.

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Patterns of selection against centrosome amplification in human cell lines

Marco António Dias Louro[1]*, **Mónica Bettencourt-Dias**[1], **Claudia Bank**[1,2]

1 Instituto Gulbenkian de Ciência, Oeiras, Portugal, **2** Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

* mlouro@igc.gulbenkian.pt

OPEN ACCESS

Citation: Dias Louro MA, Bettencourt-Dias M, Bank C (2021) Patterns of selection against centrosome amplification in human cell lines. PLoS Comput Biol 17(5): e1008765. https://doi.org/10.1371/journal.pcbi.1008765

Editor: Jacopo Grilli, Abdus Salam International Centre for Theoretical Physics, ITALY

Received: March 6, 2020

Accepted: February 3, 2021

Published: May 12, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pcbi.1008765

Data Availability Statement: The code developed and used to produce the results of this manuscript can be found in https://gitlab.com/madlouro/centriole-number-variability. All other relevant data

Abstract

The presence of extra centrioles, termed centrosome amplification, is a hallmark of cancer. The distribution of centriole numbers within a cancer cell population appears to be at an equilibrium maintained by centriole overproduction and selection, reminiscent of mutation-selection balance. It is unknown to date if the interaction between centriole overproduction and selection can quantitatively explain the intra- and inter-population heterogeneity in centriole numbers. Here, we define mutation-selection-like models and employ a model selection approach to infer patterns of centriole overproduction and selection in a diverse panel of human cell lines. Surprisingly, we infer strong and uniform selection against any number of extra centrioles in most cell lines. Finally we assess the accuracy and precision of our inference method and find that it increases non-linearly as a function of the number of sampled cells. We discuss the biological implications of our results and how our methodology can inform future experiments.

Author summary

Human cells possess small structures called centrioles, which need to be duplicated and properly segregated to ensure cell viability. Paradoxically, cells with a variable number of excess centrioles are commonly found in cancer. It is thought that these cells arise from centriole overproduction and are subsequently eliminated by selection, such that their frequency is stable in the population. However, it is not known if this overproduction-selection balance is sufficient to explain the observed intra- and inter-population variation in centriole numbers.

In this study, we model the cell population dynamics of abnormal centriole numbers inspired by classical evolutionary theory, and infer overproduction and selection parameters from a panel of 67 human cell lines. Surprisingly, our results indicate that the observed variability in most cell lines can be best explained by models assuming a single penalty against extra centrioles, regardless of their number, and complex overproduction "rules", where multiple centrioles can be gained in a single event. Furthermore, we estimate that selection against extra centrioles is generally very strong. Our work presents a novel quantitative approach to analyse centriole number variation and to further our understanding of the role of centriole number abnormalities in cancer development.

are within the manuscript and its Supporting information files.

Funding: This work was supported by EMBO Installation Grant IG4152 and by ERC Starting Grant 804569 – FIT2GO, and FCT (Fundação para a Ciência e Tecnologia) grant PTDC/BIA-BID/32225/2017. M.A.D.L. was funded by FCT research fellowship PD/BD/139217/2018. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Centrioles are microtubule-based structures that organise the centrosome and thereby orchestrate microtubule nucleation in vertebrate cells [1, 2]. Centriole number abnormalities are a source of phenotypic heterogeneity in cancer cells. Indeed, centriole numbers show variability both within cancer cell populations and between different cancers [3–5]. The causes and consequences of this heterogeneity are still poorly understood. Moreover, to the best of our knowledge, there exists no quantitative description of how centriole numbers are distributed in cancer cells.

In a proliferating cell population, cells start the cell cycle with two centrioles, which duplicate once and only once during S-phase. After cytokinesis, both daughter cells inherit two centrioles each. This centriole duplication and segregation cycle ensures that centriole number is kept constant across generations [3, 6, 7].

In stark contrast with most proliferating cells, centriole numbers are often de-regulated during cancer development. In particular, cells with abnormally high numbers of centrioles are common in tumors and cancer-derived cell lines, and have been recently identified in pre-neoplastic tissues [3, 8–12]. Interestingly, within a single population of cancer cells, individual cells often carry different numbers of centrioles. However, the number of centrioles per cell in the population seems to display a specific distribution depending on the cell type [9–12]. The source of this variability within and between cell populations is still poorly understood and calls for the development of quantitative approaches.

The occurrence of extra centrioles, termed centrosome amplification, tends to bear deleterious consequences for the cell by triggering multipolar divisions, cell cycle arrest, and/or by promoting chromosome missegregation [13–16]. Thus, excess centrioles are typically counter-selected and rarely observed in healthy tissues. However, some mechanisms are known to provide protection against centrosome amplification. For example, centrosome clustering mechanisms allow cells to group extra centrioles in two spindle poles, thus improving the viability of daughter cells [13, 14, 16]. Thus, cancer cell lines are generally regarded as being more tolerant to centrosome amplification than normal cells.

Recent data suggest that centriole numbers are maintained at an equilibrium in cell line populations. For instance, it has been observed that after transient centriole elimination, p53-deficient cell populations can seemingly recover their initial distribution of centriole numbers [17, 18]. Similarly, there are reports of extra centrioles being lost over time in cell populations after induction of cytokinesis failure [19, 20]. Since centrosome amplification is typically deleterious for cells, it is likely that centriole numbers in these populations are maintained by a balance of centriole (over)production and negative selection. These dynamics are similar to an evolutionary mutation-selection process, where the *de novo* appearance of deleterious variants in a population is counteracted by natural selection, eventually converging to so-called mutation-selection balance [21].

The dynamics of centriole (over)production and selection are currently unresolved. For instance, extra centrioles may be gained "smoothly" in a dose-dependent fashion through overexpression of key centriole biogenesis regulators, such as Plk4, STIL, and SAS-6 [22–25]. Alternatively, extra centrioles may be gained in sharp transitions—if an otherwise normal cell undergoes cytokinesis failure, it may restart the cell cycle with at least double the normal number of centrioles (as suggested in [26]). Similarly, it is not known if selection strength varies with the number of centrioles. For example, it is possible that centriole clustering is less efficient in resolving multipolar spindles if the cell contains a high number of extra centrioles. In the absence of protective mechanisms, it is possible that the presence of extra centrioles is deleterious, regardless of absolute centriole numbers. Thus, it is not a trivial question how centriole (over)production and selection can generate equilibrium distribution of centriole numbers,

and if these two processes are sufficient to explain the observed centriole number heterogeneity within and between cell populations.

Here, we develop mathematical models of centriole overproduction and selection against centrosome amplification that are predicated on different assumptions on how supernumerary centrioles are produced and how selection operates. We use these models to analyse recently published data on representative cell lines of the progression from Barrett's esophagus to gastroesophageal adenocarcinoma [9] and the NCI-60 panel of cancer cell lines [10]. These two data sets provide us with the opportunity to study how centriole number distributions vary along cancer progression, from pre-malignant to malignant stages, in the case of the Barrett's esophagus data set, and between different cancer types, in the NCI-60 data set.

Employing a model selection approach, we found that models featuring a constant cost of centrosome amplification, irrespective of the number of centrioles in a cell, best explain the empirical distributions for most of the cell lines. Moreover, our results suggest that the distribution of centriole numbers is generally super-exponential, which could be indicative of multi-step centriole number increments. We identified a general trend in the parameter estimates indicating strong selection against extra centrioles but we did not detect significant differences between cell lines. Using simulations, we show that our parameter estimation method is accurate and we predict that its precision increases non-linearly with the number of sampled cells. In summary, our work presents the first quantitative description of how centriole numbers evolve in proliferating cell populations with persistent centrosome amplification and provides a statistical tool for further dissecting the processes that generate within- and between-population variation in centriole numbers.

Results

A model of centriole number dynamics in proliferating cell populations

To study how centriole number distributions in proliferating cell populations are generated, we developed a general mathematical model grounded in mutation-selection theory. Our subject of focus is a population of proliferating cells subject to centriole overproduction and selection against extra centrioles. For the purpose of data analysis, we consider that individual cells are in mitosis and fully characterised by their number of extra centrioles, i, which can range from zero to an arbitrarily high upper bound, $i_{\max}$. Thus, the population can be split by centriole numbers into subpopulations such that, for example, the zeroth subpopulation ($i = 0$) represents all cells containing wild-type centriole numbers (four centrioles). Each subpopulation i has an intrinsic growth rate r_i. Centriole overproduction occurs at a rate $\mu_{i,j}$ from subpopulation i to subpopulation j, where $j > i$. Thus, we assume that there is no loss of centrioles across cell division (which would be given by a transition from i to j where $j < i$). Centriole overproduction events can be interpreted as gain of $j - i$ centrioles. Since in these cell lines there is a net increase in the number of centrioles compared to the wild-type situation, we make the simplifying assumption that there is no loss of centrioles. Cells that contain fewer than wild-type numbers were rarely observed in the analysed data sets; for simplification purposes, we disregard them in our model and in our analysis. Finally, our model is deterministic and all subpopulation frequencies P_i are continuous variables; i.e., we assume an effectively infinite population size. Taken together, the temporal rate of change in the relative frequency of cells containing i centrioles is given by the following ordinary differential equation:

$$\frac{dP_i}{dt} = \left(r_i - \sum_{j=0}^{i_{\max}} r_j P_j(t) \right) P_i(t) + \sum_{k=0}^{i} \mu_{k,i} P_k(t) - \sum_{l=i+1}^{i_{\max}} \mu_{i,l} P_i(t) \,, \text{for all } i \leq i_{\max}. \quad (1)$$

The dynamics of the population are thus described by a system of $i_{\max} + 1$ differential equations. This model is the continuous-time equivalent of the model proposed by Moran [27], in the absence of back mutation, and included in the framework of the original quasispecies model [28–30]. As previously mentioned, several lines of evidence suggest that centriole numbers in proliferating cell populations follow a stable equilibrium distribution when unperturbed [17–20, 31]. We propose the following expression that describes a fully polymorphic equilibrium distribution, i.e. allowing cells of any subpopulation to occur (see S1 Fig for a numerical example and a test of convergence from random initial conditions; see also Materials and methods), for an arbitrary value of $i_{\max}$:

$$P_i^* = \frac{\eta_i}{\sum_{j=0}^{i_{\max}} \eta_j} \quad , P_0(0) > 0 \tag{2}$$

$$\eta_i = \prod_{j=i+1}^{i_{\max}} f(j) \left(\sum_{a_o \in A(i)} \left(\prod_{k \in a_o} f(k) \prod_{m=1}^{|\Phi(i,k)|-1} \mu_{\phi_m, \phi_{m+1}} \right) \right) \tag{3}$$

where

$$f(i) = r_0 - \sum_{j=1}^{i_{\max}} \mu_{0,j} - r_i + \sum_{n=i+1}^{i_{\max}} \mu_{i,n} \quad \text{for all } i \leq i_{\max}, \tag{4}$$

and

$$A(j) = (a_o)_{o \in \mathcal{P}(S^j)} \tag{5}$$

is the sequence containing all elements of the power set $\mathcal{P}(S^j)$. The set S^j is defined as

$$S^j = y \in \mathbb{N} : 0 < y \leq j - 1 \quad . \tag{6}$$

Finally, we define

$$\Phi(j,k) = (\phi_m)_m \in \{\{0\} \bigcup \{a_j \neg k\}\}, \tag{7}$$

with a_j corresponding to the elements of $A(j)$ from the definition in Eq (5).

This set of equations determines the equilibrium balance of centriole overproduction and selection in a cell population, i.e., it determines the predicted proportion of cells with i extra centrioles in an unperturbed cell population, given an arbitrary set of overproduction rates and fitness functions. This general solution allows for the computation of analytical expressions for the equilibrium distributions and their (log-)likelihood under more specific centriole overproduction and selection scenarios, as shown below.

Distributions of centriole numbers in cell populations tend to be heavy-tailed

Our goal is to infer the balance between selection and centriole overproduction from the shape of the distribution of centriole numbers in samples from 67 cell lines [9, 10]. In these data sets, between 35 and 82 mitotic cells were sampled from a population of cultured cells, and centrioles were identified and counted by co-immunostaining of two centriolar markers.

As a first step, we characterised which type of distribution most likely underlies these data; this is helfpful for identifying more specific models for parameter inference. For example, consider a simple model where extra centrioles are produced at a constant rate μ and extra centrioles induce a uniform reduced growth rate r. This model can be obtained in our general

framework by substituting $\mu_{i,j} = \mu$ for $j = i+1$ and 0 otherwise, and $r_i = r$. Under these assumptions, Eq (2) can be written as

$$P_i^* = \frac{(1-r-\mu)\mu^i}{(1-r)^{i+1}}, \quad P_0(0) > 0 \tag{8}$$

It can be readily seen by substitution that this equilibrium is formally equivalent to the geometric distribution:

$$P(X = i) = (1-p)^i p, \tag{9}$$

where p represents the probability of success in i trials (roughly, the probability of observing an extra centriole in our model). If an analogue to the classical mutation-selection model is sufficient to explain the data, then the data should be geometrically distributed. In contrast, we observed an overrepresentation of cells with high centriole numbers (S2 Fig) compared to a geometric distribution. This is a coarse indication that the distribution of centriole numbers in cell populations is heavy-tailed. If that is the case, then more complex centriole overproduction and/or selection dynamics are required to explain the data.

To test this at the level of individual empirical distributions for each of the 67 cell lines, we fitted geometric and Waring-Yule distributions to each of the empirical population-level distributions of centriole numbers. Here, the Waring-Yule distribution represents a generalised discrete distribution that can potentially account for heavy tails. Then, we calculated the value of the X^2 statistic as a measure of goodness-of-fit for both distributions. Finally, we determined the difference between the value of the X^2 statistic, ΔX^2, under the geometric and Waring-Yule models, such that positive values indicate a better fit of the Waring-Yule distribution. Conversely, if the geometric distribution is a better fit, we expect that both distributions should converge and yield a ΔX^2 of approximately 0.

Visual inspection of model fits suggests the Waring-Yule (heavy-tailed) distribution is a better fit to the represented empirical distributions (Fig 1A–1C). In addition, our results indicate positive values of ΔX^2 for the majority of cell lines, suggesting a better fit of the Waring-Yule (heavy-tailed) distribution (Fig 1D). For 16 out of 67 cell lines, we obtained values of $\Delta X^2 \approx 0$ (more accurately, ≤ 1), indicating exponential-like and not heavy tails. Although the control cell lines used in the NCI-60 study rank in the bottom half of cell lines ordered by ascending ΔX^2 value, apart from HaCat, they are not clearly separated from the remaining cell lines. Thus, our results suggest that a simple model reminiscent of classical mutation-selection balance, yielding a geometric distribution of centriole numbers in the population, fails to explain the data for most cell lines. However, it should be noted that some of the 16 cell lines included in the group with near-zero ΔX^2 values contain few cells with centrosome amplification in the sampled population (e.g. OACP4—1 out of 61 cells with centrosome amplification; IGROV1—1 out of 58 cells with centrosome amplification), in which case there is little information to distinguish between geometric and Waring-Yule distributions.

Ultimately, we are interested in relating the distribution of centriole numbers in the population to the processes that generate it. If the distributions are indeed heavy-tailed, this could be achieved either by weak selection against cells with high centriole numbers or by more complex centriole overproduction mechanisms. For example, since centrioles duplicate in most healthy cells, it is possible that centriole overproduction also occurs in multiples of two (which we will refer to as overduplication). Similarly, after cytokinesis failure, a cell may restart the cell cycle and reduplicate all four centrioles, gaining four extra [26]. Intuitively, overduplication or cytokinesis failure could produce cells with multiples of two and/or four centrioles.

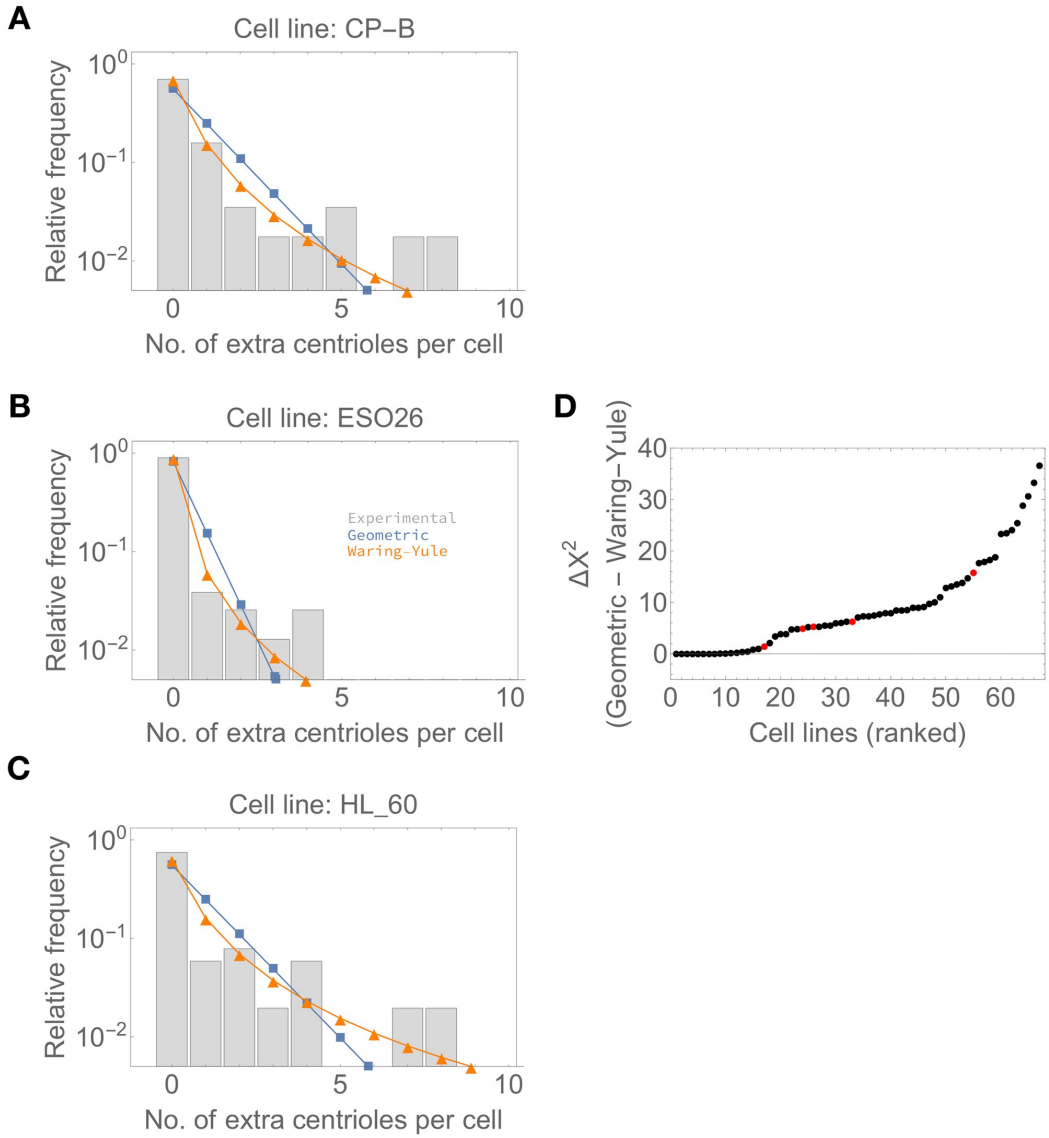


Fig 1. Distributions of centriole numbers in cell populations tend to be heavy-tailed. A-C—Examples of empirical distributions for three cell lines (grey bars) and the predicted relative frequencies under geometric (blue) and Waring-Yule distributions (orange), which are representative of distributions with exponential-like and heavy tails, respectively. Number of sampled cells: (A) $n = 57$; (B) $n = 78$; (C) $n = 51$. D—Difference between the calculated X^2 value under geometric and Waring-Yule distributions. Cell lines were ranked in ascending order according to the ΔX^2 value. The control cell lines used in [10] are highlighted in red. Higher positive values indicate a better fit of the Waring-Yule distribution to the corresponding empirical distribution, suggesting heavier-than-exponential tails.

https://doi.org/10.1371/journal.pcbi.1008765.g001

Coherently, it can be seen in both individual and pooled distributions that cells with four and eight extra centrioles are particularly abundant in the data (S2 Fig).

In summary, a model parameterised by a single centriole overproduction rate and a single intrinsic growth rate of cells with extra centrioles is not sufficient to explain the data. By visually inspecting the empirical distributions, we hypothesised that including more complex centriole overproduction events in the models, such as centriole overduplication, may provide better fits.

A candidate set of models based on centriole biology

The general model described in (1) provides a powerful starting point for inferring the distribution of centriole numbers in proliferating cell populations but it is overparameterised with respect to the data under consideration. Moreover, we are interested in comparing different hypotheses regarding specific fitness functions and overproduction parameters that could generate said centriole number distributions. To avoid overfitting and to inspect relevant biological scenarios, we generated 12 candidate models by imposing a set of constraints on centriole overproduction and on selection, based on the previous analysis of the tail of the distributions.

Several cellular processes are known to yield supernumerary centrioles [19, 20, 22–25] but we still lack a quantitative description of their contribution to the generation of cells with extra centrioles. We reasoned that these processes may be parameterised as the rate of gain of a given number of centrioles. As a universal scenario across all models, we considered that extra centrioles can be gained one at a time, at rate μ_1. Since centrioles typically duplicate in number and since we observed an excess of multiples of two and four centrioles in the data, we reasoned that centriole overproduction could also be thought of in terms of extra duplication events. Thus, we considered two additional overproduction "rules", which state two and four extra centrioles can be gained at rates μ_2 and μ_4, respectively.

Similarly, how strongly selection acts depending on the number of extra centrioles remains unknown. We focused on two main possibilities: first, that any abnormal number of centrioles is equally deleterious (resulting in a flat fitness function for all $i > 0$), and second, that the deleterious effect of extra centrioles increases with their number in a cell. Regarding the latter, we assume either an additive or power-law relationship between the number of extra centrioles and intrinsic population growth rate. In all models, we set the intrinsic growth rate for cells with wild-type centriole numbers, r_0, to be maximal and equal to 1; i.e., cells that contain no excess centrioles always have maximum fitness. Otherwise, a fully polymorphic equilibrium (i.e. where cells with any number of extra centrioles could, in theory, be observed) would not be reached under these models. The intrinsic growth rates for cells with abnormal centriole numbers ($i > 0$) are defined as (1) $r_i = r$ for all $i > 0$ ("flat" model), (2) $r_i = 1 - \frac{c \cdot i}{i_{\max}}$ for all $i > 0$ ("linear" model), and (3) $r_i = 1 - \frac{\log(i+1)}{\lambda \log(i_{\max}+1)}$ for all $i > 1$ ("power-law" model). In effect, all these functions represent a fitness landscape in which wild-type-like cells reside on a single fitness peak and all cells with extra centrioles suffer some form of fitness penalty.

The full combinatorial set of the above-fitness functions and "overproduction" rules yields 12 different models, in the following named using the initial of the fitness function and the overproduction steps (Fig 2). For example, F1- - refers to the model featuring a flat fitness function (parameterised by r) and single centriole overproduction events (parameterised by μ_1). Note that the models with fewer parameters are nested within the more parameter-heavy models and can be obtained by setting excluded centriole overproduction rates to zero. For instance, F124 yields identical equilibria to F1- - if $\mu_2 = 0$ and $\mu_4 = 0$.

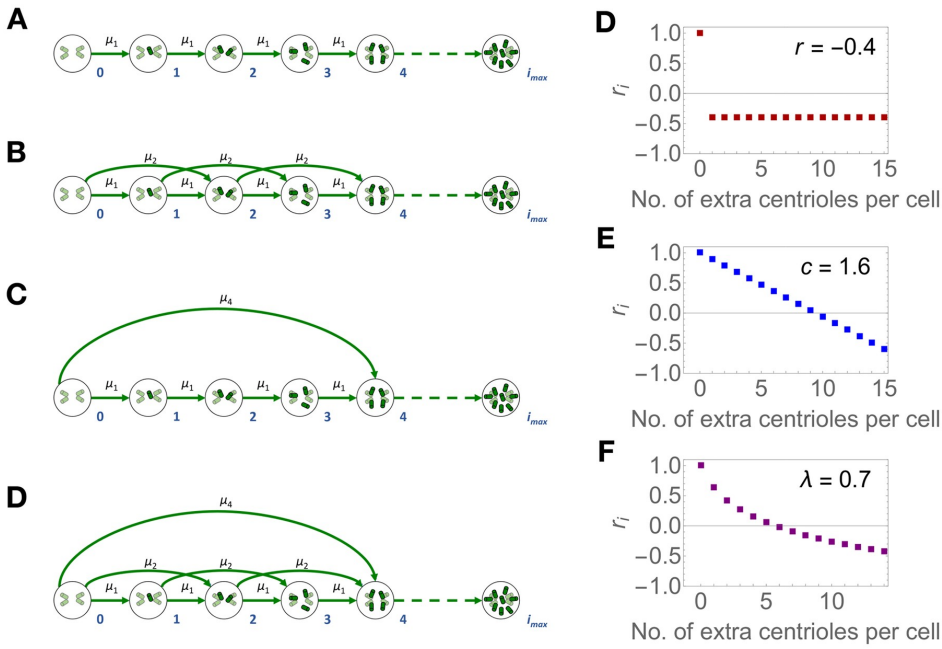


Fig 2. Centriole overproduction and fitness functions in the candidate models. A—Single-step centriole overproduction at rate μ_1. B—single- and double-step centriole overproduction events at rates μ_1 and μ_2, C—single- and quadruple-step centriole overproduction events at rates μ_1 and μ_4, D—and all three centriole overproduction events. Circles represent the subpopulation of cells containing i extra centrioles (green cylinders, i indicated by the numbers in blue). Green arrows represent transitions between subpopulations, which correspond to centriole overproduction, occurring at rates μ_1, μ_2 or μ_4. Overproduction events can occur for all i up to i_{max}. (E-G) The value of r_i for cells with i centrioles under the (E) flat, (F) linear, and (G) power-law fitness functions, evaluated at the indicated parameter values.

https://doi.org/10.1371/journal.pcbi.1008765.g002

In addition, the 12 models in our candidate set can be obtained by modifying Eq (2) according to the specified centriole overproduction and intrinsic growth rate "rules". In the case of model F124 and by extension all models nested in it simplify to a more concise form that is independent of i_{max} (see supplementary Mathematica Notebook), therefore bypassing the need to assume a potentially artificial upper bound for the number of centrioles per cell.

We note that our focal set of models is neither an exhaustive nor systematic exploration of all possibilities. However, it includes models of different complexity, depending on the number of centriole overproduction parameters. Moreover, it incorporates different biological hypotheses with respect to fitness, and results in both exponential and heavy-tailed equilibrium distributions as described above (see also S3 Fig).

Models assuming a flat fitness function best explain the data for most cell lines

As a first approach we tested if the most complex models in our candidate set (i.e. the models assuming all three overproduction events) are a good fit to the data. First, we fitted the models to each empirical distribution. Then, we performed a Monte Carlo multinomial test to distinguish between predicted and empirical distributions (Fig 3). We considered that the models were a poor fit if the test yielded a significant p-value (p-value $\leq$0.05/67, adjusted according to

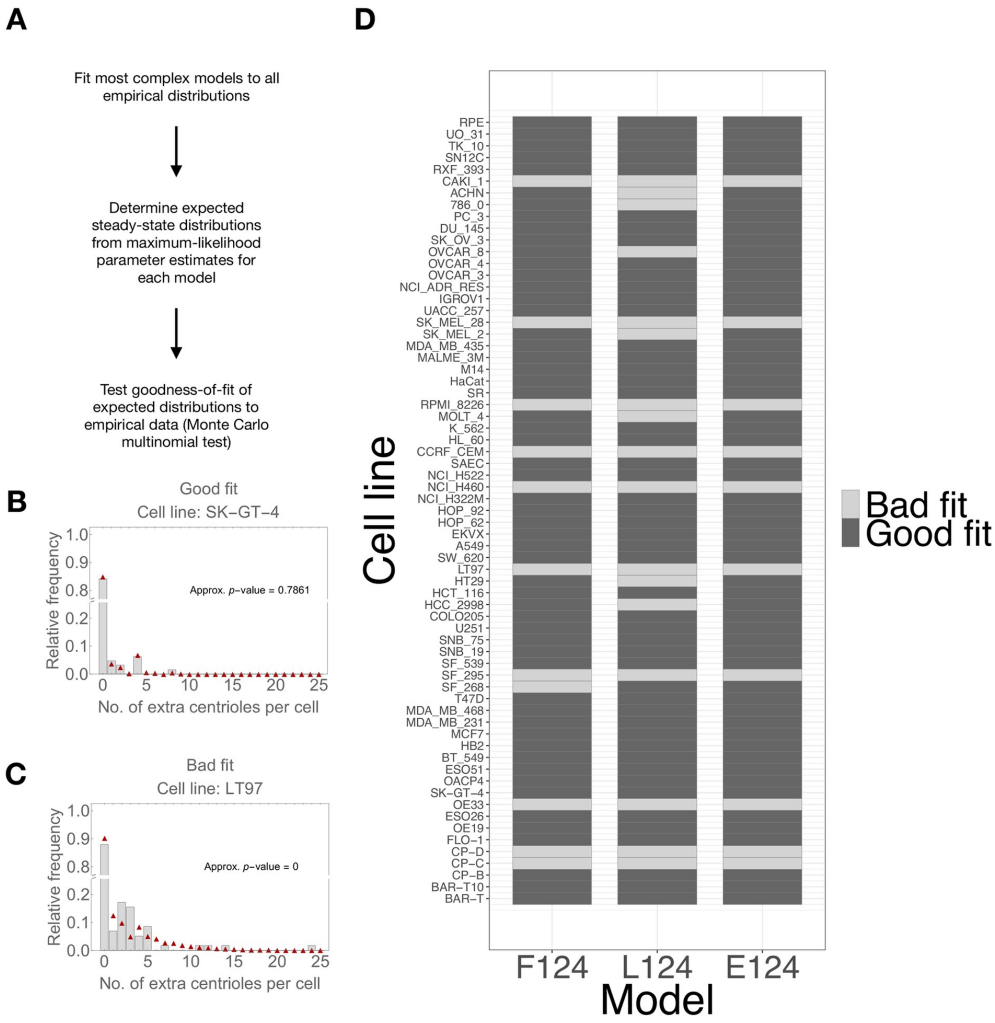


Fig 3. The most complex models are a good fit to the majority of empirical distributions. A—Procedure for determining goodness-of-fit. The significance value was set to $\alpha = 0.05$, and adjusted according to the Bonferroni correction for 67 tests. Note that under the null hypothesis, the predicted distribution under a given model is identical to the empirical distribution. B-C—Experimentally observed frequencies of centriole numbers per cell (grey bars) and the predicted frequencies under model F124—flat fitness function, single-, double-, and quadruple-step overproduction parameters (red triangles). The two examples include cases in which we obtained non-significant (good fit) and significant (bad fit) p-values. Number of sampled cells: B—$n = 82$. C—$n = 63$. D—Goodness-of-fit of the three most complex models for each cell line.

https://doi.org/10.1371/journal.pcbi.1008765.g003

a Bonferroni correction). Our models showed a good fit for a majority (56 for the "flat" model, 52 for the "linear" model, 57 for the "power-law" model) of cell lines.

Thus, we concluded that the most complex models are a good fit to the data. For 10 out of 67 cell lines, all three models were a poor fit to the corresponding empirical distributions. These cell lines tended to display some proportion of cells with extremely high centriole

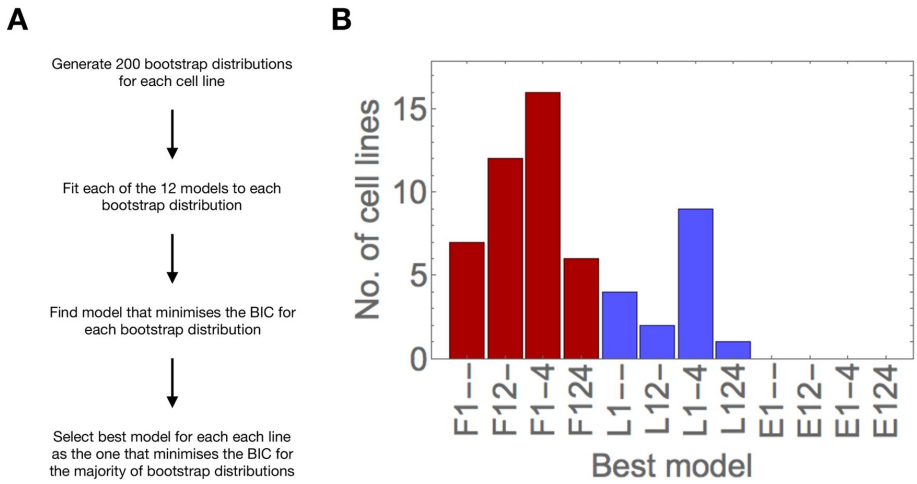


Fig 4. Models assuming a flat fitness function best explain the data for most cell lines. A—Procedure for model selection. B—Models with a flat fitness function explain the data for most cell lines. Number of cell lines for which the corresponding model was selected as the best. The fitness function of the models is indicated in red (flat) or blue (linear).

https://doi.org/10.1371/journal.pcbi.1008765.g004

numbers (≥15), which are very rare under all our models. We took a conservative approach and excluded these cell lines from further analysis.

Next, we asked which of the 12 models best explains the observed data. We generated 200 bootstrap distributions for each cell line by drawing a random sample with replacement from each empirical distribution, and fitted each of the 12 models. Then, we calculated the Bayesian Information Criterion (BIC) score from the resulting maximum log-likelihood value. The model that minimised the BIC for the largest number of bootstrap distributions was selected as the best for each cell line (Fig 4). Strikingly, the best models for 41 out of 57 cell lines assumed the flat fitness function, including all six lung and kidney cell lines in the NCI-60 panel, and the two metaplasia and one dysplasia cell lines in the Barrett's esophagus data set (see S4 Fig for model selection results grouped by tissue of origin). 16 cell lines are best explained by models assuming the linear fitness function; these cell lines are not associated to any specific tissue types or developmental stages. No cell line is best explained by models with a power-law fitness function. In contrast, there is more variability with respect to the best set of centriole overproduction parameters.

Next we analysed the number of bootstrap distributions that was selected for each cell line (S5 Fig). We observed that the best models for each cell line were selected for a maximum of 198 (99%) and a minimum of 40 (20%) bootstrap distributions, with a median of 110.5 (55.25%) bootstrap distributions (S5(B) Fig). In addition, in some cases, the BIC score was equal for models assuming the flat and power-law fitness functions (see supporting code). This means that the decision for the most appropriate model is sometimes not clear, which can be either due to the models yielding indistinguishable distributions, or the data not being sufficiently informative to distinguish between models (see also below).

Then we examined models sharing the same fitness function as the best model. In total, the same fitness function was selected for a maximum of 199 (99.5%) and a minimum of 78 bootstrap distributions (39%), with a median of 150 (75%) (S5(C) Fig). Therefore, the same fitness function was selected more consistently than individual models.

Despite some uncertainty in model selection, we concluded that most empirical distributions are best explained by models assuming a flat fitness function, whereas a smaller subset of cell lines was best explained by models with a linear fitness function.

Parameter estimates indicate strong selection against excess centrioles

We next tested if we could distinguish between different cell lines based on their parameter estimates. Following our model selection results, we focused on model F124, which includes a constant fitness function and allows for all possible overproduction rules. First, we tested the sensitivity of the fitting as a function of the parameters by evaluating the log-likelihood expression at values around the maximum, and looked for correlations between pairs of parameters. We observed that all parameters are uncorrelated for model F124, and thus can be independently estimated from the data (S6 Fig).

Second, we analysed the maximum-likelihood r and μ_1 estimates obtained for the previously generated non-parametric bootstrap distributions, under model F124 (Fig 5). We observed globally negative median estimates of the intrinsic growth rate r, indicating strong selection against cells with extra centrioles. The median estimates for the single-step overproduction rate μ_1 were relatively low but more variable than those of the intrinsic growth rate r. Indeed, for most cell lines, the median estimate of μ_1 fell within the range of 0-0.423, with two cell lines scoring over 0.7 (SE268 and HT19). However, the confidence intervals for both parameters were considerably wide, spanning almost the whole parameter range in the case of the intrinsic growth rate r, such that we could not identify significant differences between cell lines or by tissue of origin. Thus, our results indicate a pattern of low to moderate centriole overproduction rates and intrinsic growth rates for cells with extra centrioles.

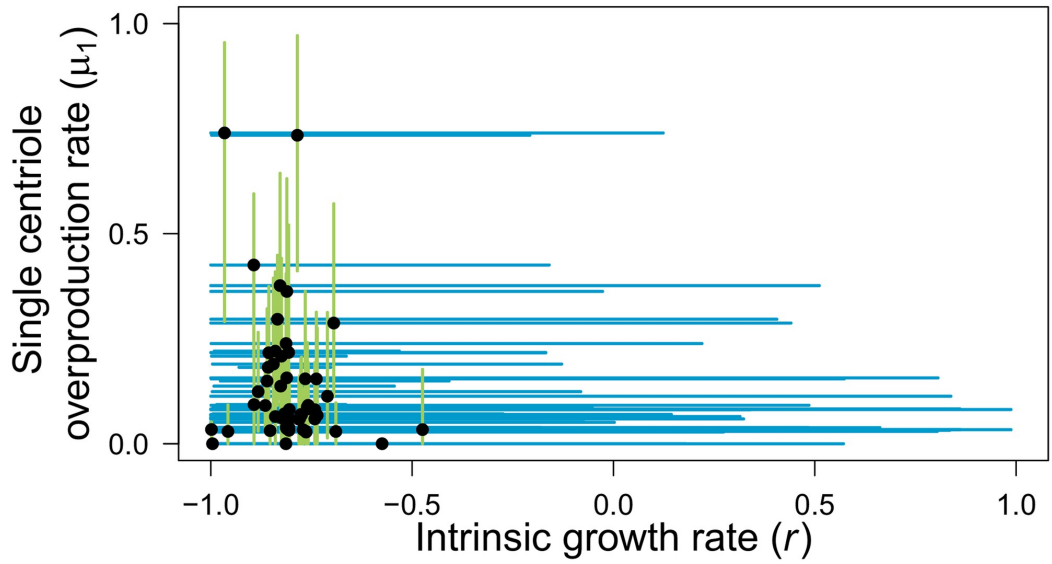


Fig 5. Parameter estimates indicate strong selection against excess centrioles but there is considerable uncertainty in parameter estimation. Median estimates (black dots) of r and μ_1 obtained from 200 non-parametric bootstrap distributions for each empirical distribution. Lines indicate 95% confidence intervals (0.025 to 0.975 inter-quantile distance) for the intrinsic growth rate r (blue) and the single-step overproduction rate μ_1 (light green).

https://doi.org/10.1371/journal.pcbi.1008765.g005

Next, we addressed the accuracy and precision of our inference method. If our method is accurate, parameter estimates from model simulations should converge to the input parameter values. In addition, we expected that errors in parameter estimation should be similar between model simulations and the ones obtained in the non-parametric bootstrap, for samples of the same size. To test this, we performed parametric bootstrapping, where we resampled the expected distributions under model F124 instead of resampling the empirical distributions as in the non-parametric bootstrap. As input values, we used the previously calculated median r, μ_1, μ_2, and μ_4 values from the 200 non-parametric bootstrap distributions for each cell line. We assumed a sample size, i.e. number of sampled cells, equal to that obtained in the corresponding data set and generated 200 parametric bootstrap distributions. Then, we fitted model F124 to each parametric distribution and analysed the estimated parameter values.

Our results show that the median parameter estimate obtained from the parametric mostly agrees with the input value (S7 Fig), indicating that our method is accurate. Moreover, confidence interval lengths were similar to those obtained from the non-parametric bootstrap distributions. For some cell lines, the errors obtained for the intrinsic growth rate r differed between the non-parametric and parametric bootstrap distributions. This is likely due to the lower sensitivity of the maximum likelihood values to changes in r compared to μ_1 (S6 Fig).

We concluded that parameter estimation accuracy and precision are similar when data is simulated either from the empirical or predicted distributions. Since the confidence interval length is influenced by sample size and we did not detect systematic biases in our inference method, it is likely that the precision of our parameter estimates mainly depended on the number of sampled cells per cell line.

Accuracy and precision of the inference method increase non-linearly with the number of sampled cells

To provide statistical guidance for future experiments, we asked how much the precision of the parameter estimates and accuracy of model selection could be improved by increasing the sample size, within experimentally feasible limits. To address this question, we reasoned that the estimated values from the median expected distribution would provide a useful test case. We identified the median expected distribution by measuring the Euclidean distance between the vector whose elements are the four estimated parameter values for a given bootstrap distribution to the vector of the lowest possible values for each parameter (equating to [−1, 0, 0, 0] for r, μ_1, μ_2, and μ_4, respectively).

We used the parameter values corresponding to the median Euclidean distance as input for model F124 and simulated 200 parametric bootstrap distributions in a range of sample sizes. For comparison purposes, we assumed a sample size of 35 and 83, which correspond respectively to the lowest and the highest number of cells obtained experimentally in our data sets. In addition, we considered realistic sample sizes of 50, 100, 150, 200, and 250 cells. Finally, we simulated parametric bootstrap distributions with a sample size of 1000 to analyse the properties of our inference method if larger sample sizes were attainable.

Subsequently, we fitted the three most complex models. We counted the number of times each model maximised the log-likelihood function out of the 200 bootstrap distributions for each simulated sample size. Note that the BIC is unnecessary because the three models have the same number of parameters. In addition, we analysed the parameter estimates of model F124 from the simulated distributions.

Since we performed simulations under model F124, this model should fit best most of the bootstrap distributions. It is also trivial that errors in parameter estimation will become smaller

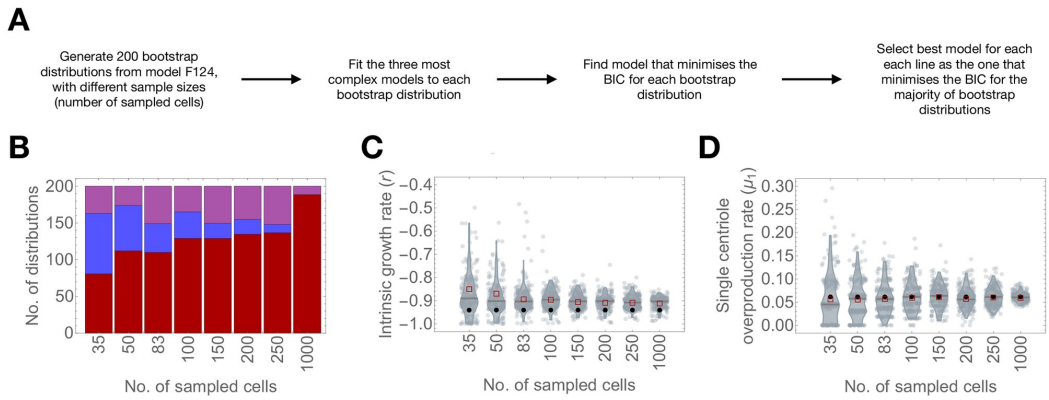


Fig 6. Accuracy and precision of the inference method increase non-linearly with the number of sampled cells. A—Number of bootstrap distributions for which model F124 (red), L124 (blue), and E124 (purple) fitness functions were selected as the best in simulations of model F124 as a function of bootstrap sample size. Input parameter values for the simulations: $r = -0.940$, $\mu_1 = 0.062$, $\mu_2 = 0.096$, $\mu_4 = 0.288$. B-C—Bootstrap distribution of parameter estimates for intrinsic growth rate r (B) and single-step overproduction rate μ_1 (C) as a function of bootstrap sample size. Black dots indicate the median estimated value and the green dot indicates the input value used in the simulations.

https://doi.org/10.1371/journal.pcbi.1008765.g006

for larger sample sizes. Regardless, we were interested in quantifying how often model F124 is identified as the best model and how confidence interval length varies with the number of sampled cells.

Fig 6 shows the results for model selection from simulated data. We observed that within the range of experimentally obtained sample sizes, model F124 is correctly identified in 81 (40.5%), 112 (56%), and 110 (55%), out of 200 parametric bootstrap distributions, for a sample size of 35, 50, and 83 simulated cells, respectively. Model F124 is only marginally outperformed by model L124 for a sample size of 35 (selected for 81 and 82 bootstrap distributions, respectively). Nevertheless, we observed that the number of bootstrap distributions for which the true model is selected increases to 129 (59.5%) for a sample size of 100, and to 137 for a sample size of 250 (67.5%). For the maximum sample size tested, the true model was selected for 189 (94.5%) bootstrap distributions.

Next, we inspected how the distributions of parameter estimates vary with sample size. We observed that the input value always falls within the confidence interval but we noted that the median intrinsic growth rate r was consistently overestimated with respect to the input value (Fig 6C). Nevertheless, parameter estimation was fairly accurate for higher sample sizes. We also observed a near two-fold decrease in confidence interval length between the minimum (35) and maximum (83) experimentally obtained sample sizes, and a further 1.55, 1.91, 2.02 and 2.01-fold decrease between the maximum experimentally obtained sample size (83) and examples containing 100, 150, 200, and 250 simulated cells, respectively. For 1000 simulated cells, we obtained confidence intervals with a length of 0.109, corresponding to a 2.43-fold decrease compared to the maximum experimentally obtained sample size.

Unlike for intrinsic growth rates, we obtained extremely accurate median estimates for μ_1 regardless of sample size (Fig 6D). Thus, it is possible that the overestimation of the intrinsic growth rate r is because the estimated values lie close to the lower bound (-1), and not an inconsistency generated by the model. We observed that the confidence interval for the single-step overproduction rate μ_1 decreased from 0.194 to 0.138 (1.41-fold decrease), for sample sizes of 35 and 83, respectively, and further to 0.114 for a sample size of 100 (1.21-fold decrease

compared to a sample size of 83) and 0.083 for a sample size of 250 (1.66-fold decrease compared to a sample size of 83). For 1000 simulated cells, confidence interval lengths were as low as 0.0412.

In conclusion, both model selection accuracy and parameter estimation precision increased non-linearly with sample size. Importantly, we predict that increasing the number of sampled cells within a feasible range (between 100 and 200) can greatly improve our inference power. However, it should be stated that these results may change depending on the number of cells with centrosome amplification. For example, if there are few abnormal cells in the population, it is probably harder to distinguish between models. Conversely, if cells with extra centrioles are more frequent, it is expected that models become easier to distinguish. In addition, the range of parameter values should be taken into account to avoid estimation biases, such that estimated values do not fall close to the bounds.

Discussion

We here combined analytical and statistical methods to characterise abnormal centriole number distributions in populations of human cell lines. Adopting classical mutation-selection balance theory from population genetics, we developed a set of mathematical models for analysing a broad panel of cell lines, which are representative of the diversity along cancer development and between different cancer types. Using a model selection approach, we found that a constant and heavy cost of excess centriole numbers is a common feature of the best approximating models for the majority of cell lines. In addition, we quantified how uncertainty in the model selection and parameter estimation procedures can be reduced by obtaining larger sample sizes in the future. We show that integrating statistical information into experimental setups could reveal potential differences between cell lines in the mechanisms that cause abnormal centriole number distributions. Importantly, our population-level approach recognises and quantifies the variation in centriole numbers that has recently been observed in experimental data.

Dynamics of centriole numbers in proliferating cell populations

To the best of our knowledge, we provide the first quantitative description of centriole number dynamics in populations of proliferating cells. Past studies, both experimental and theoretical, on the population-level response to supernumerary centrioles investigated how wild-type numbers are recovered after perturbation [17, 18], and also highlighted the role of negative selection in driving this process. These previous studies have mainly focused on distinguishing cells with wild-type centriole/centrosome numbers from cells with abnormal numbers. Here, we explored the full centriole number variation that has been observed in experimental studies. We described the heterogeneity in centriole numbers per cell within and between cell populations, and we evaluated which type of underlying fitness function is most likely to generate the observed variation within the population. Interestingly, our analysis suggests that selection acts strongly against any number of excess centrioles in most cell lines. This means that deleterious effects arise as soon as excess centrioles are produced, whereas the actual number does not seem to matter for selection, and that the shape of the distribution is determined chiefly by the mechanism(s) of centriole overproduction.

Implications for the biology of centrosomes

Our results show that the centriole number distribution within a population carries important biological information. First, as we argued above, we inferred a constant and heavy cost of abnormal centriole numbers. Whereas understanding variation has long been a staple of

evolutionary studies, it has often been overlooked in cell biology. However, on a similar subject, it has been reported that different mechanisms of organelle biosynthesis, such as *de novo* assembly or fusion, display specific signatures that can be identified by relating the mean and the variance in their distributions [32]. Thus, valuable insight can be gained from a broader quantitative description of the data.

Second, simple models incorporating single-centriole overproduction events and a constant fitness function (i.e. akin to the classical formulation of mutation-selection balance in population genetics and widely explored in quasispecies models as reviewed in [33]) were sufficient to explain the shape of the centriole distribution in a few cell lines, whereas in others a more complex relationship between selection and overproduction improved the fitting. In the latter case, our analysis indicates that the shape of the distribution depends more on how supernumerary centrioles are acquired rather than on how they are eliminated—i.e. model fits are less sensitive to changes in intrinsic growth rates than to the mode of overproduction. This raises the hypothesis that various cellular mechanisms might lead to overproduction whereas selection "punishes" the presence of any number of excess centrioles.

The biological processes associated with our postulated single-, double- and quadruple-step centriole overproduction events may be entirely different. For example, overproduction of two centrioles could occur due to centriole re-duplication after premature disengagement [34] in wild-type cells. Quadruple overproduction could be a consequence of cytokinesis failure followed by reduplication of all four centrioles [26].

We identified models with a flat fitness function as the best for most cell lines, suggesting that centrosome amplification *per se* is deleterious, regardless of the number of extra centrioles. Intuitively, one could expect that a higher number of centrioles would induce stronger selection because the mechanisms that eliminate cells with centrosome amplification seem to involve some form of "counting". For example, one of the main sources of cell death in cells with extra centrioles is multipolar divisions [13, 14]. It is possible that it is harder for cells to cluster extra centrioles if there are more of them. Thus, one could expect that the probability that a cell undergoes a multipolar division increases with the number of extra centrioles. Likewise, it has been recently proposed that a molecular complex called the PIDDosome triggers p53-dependent cell cycle arrest by "counting" excess mother centrioles [15]. Thus, it seems likely that the efficacy of the PIDDosome in detecting extra centrioles should increase with the number of extra centrioles. If the probability of multipolar divisions and PIDDosome-dependent cell cycle arrest increases with the number of extra centrioles, this could imply a linearly decreasing fitness function as the one we inferred for some cell lines. However, it is still not clear how these or other mechanisms respond to the number of extra centrioles.

We anticipate that future experimental work will address these mechanisms in greater detail, upon which our models can be refined to integrate mechanistic details of overproduction rather than the current general overproduction rates. That could, in turn, allow for a more specific statistical inference of when and how centriole overproduction occurs in different cell lines. Ideally, our modeling and inference approach will eventually link experimental information about centriole distributions with genomic inference of cancer-line and -stage specific molecular alterations.

Limitations of this study

The observation that cells with extra centrioles are relatively rare in the analysed data sets is a major determinant of the uncertainty of our model selection and parameter estimation procedures. However, we showed that a modest increase in the number of analysed cells can potentially mitigate these issues. In addition, mapping model parameters to the biological system

may not be trivial. Importantly, one of the aspects we simplified is how extra centrioles are allocated to the daughter cells after mitosis, or centriole segregation. Whereas after correct centriole duplication, each daughter cell inherits a single centriolar pair, it is still unclear how extra centrioles are distributed. However, while this manuscript was in review, Sala et al. reported that the degree of asymmetry in centriole segregation increases with the number of extra centrioles in RPE-1 cells transiently overexpressing Plk4 [31]. Since most cells with centrosome amplification in the samples we have analysed contain only a few extra centrioles, it can be expected that centriole segregation is predominantly symmetrical. Moreover, this study reports that extra centrioles are capable of duplication. Thus, cells should have the same number of centrioles as their mothers on average. In light of these results, we conclude that our deterministic approximation is reasonable.

Furthermore, the fact that cells in the data sets are mitotic is convenient for the inference of centriole overproduction and selection parameters because allowed us to eliminate confounding variables related to the cell cycle. First, it means that centriole number variations along the cell cycle, which would be expected in an asynchronously dividing population, need not be considered. Second, it allowed us to disregard differences in cell cycle progression within and between cell lines. Altogether, characterising centriole number distributions along cell cycle progression is an entirely different problem. From a purely theoretical standpoint, it requires a more detailed implementation of the timing of centriole duplication and the length of each cell cycle phase. This would be desirable in the broader context of characterising centriole number distributions in proliferating cell populations but is beyond the scope of our current analysis.

Towards unraveling the causes and consequences of centriole number changes in cancer

Numeric aberrations of centrioles and their putative link to cancer formation have long been described [3], although accurate quantifications of centriole numbers in tumor biopsies and cancer-derived cell lines have emerged only recently [9–12]. To this day, the contribution of centrosomal anomalies to cancer development remains controversial, with some studies showing that higher numbers, via Plk4 overexpression, can initiate or aggravate tumorigenesis [35, 36], and others showing that it is not sufficient and may even slow down progression [37, 38]. On the other hand, extra centrioles are associated with other cancer hallmarks, such as aneuploidy [14, 39] and invasion [40–42], and often correlate with a more aggressive cancer phenotype [3, 7]. The widespread occurrence of centriole number abnormalities in a cancer setting makes them attractive as prospective biomarkers and as therapeutic targets [43]. Thus, understanding the underlying causes of these abnormalities is important for biomedical research.

Here, we adopted evolutionary theory to quantify the variation in centriole numbers within cancer cell populations. That is because ultimately, centriole number abnormalities are highly heterogeneous both within and between cancer cell populations. In order to predict how these abnormalities evolve during cancer development, and how they may interweave with other cancer hallmarks, it is crucial to have a quantitative understanding of how extra centrioles emerge in these cells, and how the cells cope with them. For example, in the case of the Barrett's esophagus progression model, the increase in centriole numbers from the metaplasia to the dysplasia stages can be explained by loss of p53 [9]. This can be interpreted as a reduction in the strength of negative selection, since p53 can lead to cell-cycle arrest or cell death in the presence of extra centrioles. If this was true, it would be interesting to quantify how strong the decrease in selective pressure and if it is sufficient, by itself, to account for the shift in centriole number distributions.

We showed that, pending an increase in statistical power, our modelling framework can be used to infer these changes and further our understanding of the relationship between extra centriole numbers and cancer development.

Materials and methods

Experimental data

For our analysis, we considered two recently published data sets. The first data set corresponds to 13 cell lines derived from different tissues in the Barrett's esophagus cancer progression model. These include pre-malignant (metaplasia and dysplasia) and malignant stages (adenocarcinoma and lymph node metastasis). The second data set corresponds to 53 cell lines from the NCI-60 panel, a group of cell lines that spans multiple cancer types (leukemia, melanoma and lung, colon, brain, ovary, breast, prostate, and kidney, cancers), as well as five non-cancerous cell lines. In both cases, the data correspond to centriole number counts in mitotic cells. Out of the 71 cell lines, four were discarded from the analysis: for three of the cell lines, we could not compute the equilibrium expression, due to high $i_{\max}$); for the remaining cell line, no cells with centrosome amplification were recorded, under which conditions the models can be trivially solved by setting μ_1, and/or μ_2, and/or μ_4 to zero. As previously stated, we do not take into account any cell with less than four centrioles (which represent approximately 1.9% of the total). Additional experimental details can be found in the corresponding publications [9, 10].

Mathematical analysis

Eq (1) describes the rate of change in the equilibrium frequency of the subpopulation of cells containing i centrioles. To obtain the equilibrium solution in Eq (2), we solved $\frac{dP_i}{dt} = 0$ for all i, for $i_{\max} = \{3, 4, 5\}$, using Wolfram Mathematica, and proposed a general expression for increasing $i_{\max}$. Then, we verified if the expression was correct by comparing it to the steady-state obtained from numerical integration of Eq (1), and further confirmed that the system converges to the same equilibrium for random initial conditions (S1 Fig).

To ensure Eq (2) yields exclusively non-negative values for all i, we added the following constraint:

$$\mu_{i,j} > 0 \qquad \wedge \qquad r_0 > \sum_{j=0}^{i_{\max}} \mu_{0,j} + r_i - \sum_{k=i}^{i_{\max}} \mu_{i,k}. \tag{10}$$

The first term indicates that centriole overproduction rates $\mu_{i,j}$ are strictly positive, otherwise higher centriole numbers would not be reachable. The second term indicates that the intrinsic growth rate of wild-type cells must exceed the sum of the intrinsic growth rates of cells with i extra centrioles and the rate at which their frequency increases as a function of centriole overproduction. Breaking this constraint would lead to the depletion of cells with wild-type centriole numbers, which is not observed in any of the data in our analysis.

Model fitting and selection

To fit the models, we first derived a general (log-)likelihood expression:

$$\ln \mathcal{L}(\theta_M | p_i) = \sum_{i=0}^{i_{\max}} p_i \ln (P_i^* | \theta_M) \tag{11}$$

where θ is the tuple of parameters in model M, P_i is the equilibrium solution derived in (2), p_i

Table 1. Model parameters. The indicated ranges were used as constraints when estimating parameters.

Parameter	Range
r	]−1, 1[
c	]0, 2[
λ	]0, 2[
μ_1	[0, 1[
μ_2	[0, 1[
μ_4	[0, 1[

https://doi.org/10.1371/journal.pcbi.1008765.t001

is the observed relative frequency of cells containing i centrioles, and $i_{\max}$ indicates the subpopulation of cells harboring the maximum observed number of centrioles. For ease of comparison, we assumed $i_{\max}$ to be the maximum overall number of extra centrioles per cell in the data (30). We fitted the models by numerical maximisation of the log-likelihood function, according to model parameters and the indicated range of values (Table 1).

The Bayesian Information Criterion (BIC) was calculated according to:

$$BIC = \ln(n)\kappa - 2\ln\hat{\mathcal{L}}(\theta_M|p_i), \tag{12}$$

where $\hat{\mathcal{L}}$ is the maximum likelihood estimator for a given model, κ is the number of parameters in the model and n is the sample size (number of cells observed in a given cell line). The BIC accounts for sample size and the number of model parameters, such that more complex models are penalised. When compared to another frequently used model selection criterion, the Akaike Information Criterion (AIC), the added sample size penalty is useful given that the number of sampled cells per cell line is limited. We selected the best model for each cell line by finding the one that minimises the BIC score. We performed model fitting and selection using Wolfram Mathematica. The results for model selection based on the empirical distributions alone (as opposed to the bootstrap samples) were confirmed in Python to check for numerical inconsistencies in log-likelihood values.

Bootstrapping

Non-parametric bootstrap samples were generated by drawing data points, with replacement, from the empirical distribution for each cell line. The sample size for each bootstrap distribution was equal to the one obtained experimentally. Parametric bootstrap samples were generated by drawing, with replacement, from each predicted distribution. In other words, we fitted the models to the empirical distributions in question to obtain an expected distribution under the respective models, and sampled from this distribution.

Statistical analyses and data visualisation

After obtaining expected distributions under geometric and Waring-Yule models, using Wolfram Mathematica built-in functions, we calculated the value of the X^2 statistic for each pair of empirical and expected distributions. It should be noted that the X^2 statistic has limited power for analysing the data at hand, such that we used it only for comparative purposes.

When testing goodness-of-fit of the most complex candidate models, we first determined the expected distribution under each of the models, for each cell line, and then performed Monte Carlo multinomial tests between the expected distribution and the corresponding empirical distribution. The Monte Carlo step consisted of 10,000,000 simulations under the expected distribution. We determined (approximated) p-values based on the proportions of

simulations that yielded more extreme results than the data. The significance level was set to 0.05, and adjusted according to the Bonferroni correction for multiple testing (one for each of the 67 cell lines).

All plots were produced in Wolfram Mathematica or R. Also see the supplementary code in the accompanying repository: https://gitlab.com/madlouro/centriole-number-variability.

Supporting information

S1 Fig. The solutions of the system of ordinary differential equations converge to the expected equilibrium value. A—Comparison between numeric integration of the general model and the corresponding equilibrium expression (2), evaluated at the same parameter values. We assumed $i_{max} = 15$ and generated pseudo-random parameter values for all r_i and $\mu_{i,j}$. B—Comparison between numeric integration of the general model from different initial conditions and the corresponding equilibrium expression. We generated a set of pseudo-random parameter values for all r_i and $\mu_{i,j}$ and initial conditions. Note that the y-axis is in log-scale.
(TIF)

S2 Fig. Cells with high number of centrioles occur frequently across all data sets. Best fitting geometric distribution (blue) to the pooled distribution of centriole numbers in all sampled populations. Number of sampled cells: $n = 3746$.
(TIF)

S3 Fig. Models in the candidate set can produce both geometric and heavy-tailed distributions. Best fitting geometric (blue) and Waring-Yule (orange) distributions to 1,000,000 simulated data points (relative frequencies indicated as grey bars). A—Data simulated from model F1- - ($r = -0.3$, $\mu_1 = 0.6$). B—Data simulated from model F124 ($r = -0.5$, $\mu_1 = 0.2$, $\mu_2 = 0.1$, $\mu_4 = 0.1$). Data points were generated by multinomial sampling from the equilibrium distributions evaluated at the indicated parameter values. Note that the y-axis is in log-scale.
(TIF)

S4 Fig. Best fitness functions per tissue type. Models sharing the fitness function of the best model for each cell line in the data sets. A—Barrett's esophagus data set, grouped by developmental stage. B—NCI-60 data set, grouped by tissue type (including cancer and non-cancer cell lines). "Other" refers to an RPE cell line that was used as a control.
(TIF)

S5 Fig. Bootstrap support of the candidate models. A—Number of bootstrap distributions (out of the 200 generated for each cell line) explained by each model. We obtained indistinguishable BIC scores for multiple models in 381 bootstrap distributions spread across 26 different cell lines ("Mult."). C—Number of bootstrap distributions explained by models sharing the same fitness function as the best model for each cell line. The fitness functions of the models are indicated in red (flat), blue (linear), purple (power-law), and green (multiple models).
(TIF)

S6 Fig. Model parameters are not correlated and the likelihood value shows little sensitivity to *r*. We fitted the model to a random empirical distribution from the analysed data sets and calculated the likelihood values centered around the maximum as a function of pairwise combinations of parameter values. A—r and μ_1; B—r and μ_2; C—r and μ_4; D—μ_1 and μ_2; E—μ_1 and μ_4; F—μ_2 and μ_4. Note that the scale for r is different from that of the centriole overproduction parameters μ_1,μ_2, and μ_4.
(TIF)

S7 Fig. Parameter estimation is accurate and estimation errors are likely due to small sample sizes. Difference between the median value of the parametric bootstrap distribution for each parameter value and the input value for the simulated data (in black) and difference between the confidence interval length of the non-parametric and parametric bootstrap distributions (in yellow). A—r. B—μ_1. C—μ_2. D—μ_4.
(TIF)

Acknowledgments

We would like to acknowledge Carla A. M. Lopes and Gaëlle Marteil for in-depth discussions of their work with the Barrett's esophagus and NCI-60 cell lines, and Telmo Cunha for his work on a prior version of the models. We thank all members of the Bank and Bettencourt-Dias labs for their insights and critical reading of this manuscript. We would like to acknowledge various participants of the "From Molecular Basis to Predictability and Control of Evolution" workshop at the Nordita Institute in Sweden for inspiring discussions.

Author Contributions

Conceptualization: Marco António Dias Louro, Mónica Bettencourt-Dias, Claudia Bank.

Data curation: Marco António Dias Louro.

Formal analysis: Marco António Dias Louro.

Funding acquisition: Claudia Bank.

Investigation: Marco António Dias Louro.

Methodology: Marco António Dias Louro, Claudia Bank.

Project administration: Mónica Bettencourt-Dias, Claudia Bank.

Resources: Mónica Bettencourt-Dias, Claudia Bank.

Software: Marco António Dias Louro.

Supervision: Mónica Bettencourt-Dias, Claudia Bank.

Validation: Marco António Dias Louro, Claudia Bank.

Visualization: Marco António Dias Louro.

Writing – original draft: Marco António Dias Louro, Claudia Bank.

Writing – review & editing: Marco António Dias Louro, Mónica Bettencourt-Dias, Claudia Bank.

References

1. Azimzadeh J, Marshall WF. Building the centriole. Current Biology. 2010; 20(18):R816–R825. https://doi.org/10.1016/j.cub.2010.08.010
2. Brito DA, Gouveia SM, Bettencourt-Dias M. Deconstructing the centriole: Structure and number control. Current Opinion in Cell Biology. 2012; 24(1):4–13. https://doi.org/10.1016/j.ceb.2012.01.003
3. Godinho SA, Pellman D. Causes and consequences of centrosome abnormalities in cancer. Philosophical Transactions of the Royal Society B: Biological Sciences. 2014; 369 (1650). https://doi.org/10.1098/rstb.2013.0467 PMID: 25047621
4. Bettencourt-Dias M, Hildebrandt F, Pellman D, Woods G, Godinho SA. Centrosomes and cilia in human disease. Trends in Genetics. 2011; 27(8):307–315. https://doi.org/10.1016/j.tig.2011.05.004
5. Gönczy P. Centrosomes and cancer: Revisiting a long-standing relationship. Nature Reviews Cancer. 2015; 15(11):639–652. https://doi.org/10.1038/nrc3995

6. Banterle N, Gönczy P. Centriole Biogenesis: From Identifying the Characters to Understanding the Plot. Annual Review of Cell and Developmental Biology. 2017; 33(1):23–49. https://doi.org/10.1146/annurev-cellbio-100616-060454
7. Nigg EA, Holland AJ. Once and only once: Mechanisms of centriole duplication and their deregulation in diseases. Nature Reviews Molecular Cell Biology. 2018; 19(5):297–312. https://doi.org/10.1038/nrm.2017.127
8. Chan JY. A Clinical Overview of Centrosome Amplification in Human Cancers. International journal of biological sciences. 2011; 7(8):1122. https://doi.org/10.7150/ijbs.7.1122
9. Lopes CAM, Mesquita M, Cunha AI, Cardoso J, Carapeta S, Laranjeira C, et al. Centrosome amplification arises before neoplasia and increases upon p53 loss in tumorigenesis. Journal of Cell Biology. 2018; 217(7):2353–2363. https://doi.org/10.1083/jcb.201711191 PMID: 29739803
10. Marteil G, Guerrero A, Vieira AF, De Almeida BP, Machado P, Mendonça S, et al. Over-elongation of centrioles in cancer promotes centriole amplification and chromosome missegregation. Nature Communications. 2018; 9(1). https://doi.org/10.1038/s41467-018-03641-x PMID: 29593297
11. Wang M, Knudsen BS, Nagle RB, Rogers GC, Cress AE. A method of quantifying centrosomes at the single-cell level in human normal and cancer tissue. Molecular Biology of the Cell. 2019; 30(7):811–819. https://doi.org/10.1091/mbc.E18-10-0651
12. Breslow DK, Holland AJ. Mechanism and Regulation of Centriole and Cilium Biogenesis. Annual Review of Biochemistry. 2019; 88(1):691–724. https://doi.org/10.1146/annurev-biochem-013118-111153
13. Silkworth WT, Nardi IK, Scholl LM, Cimini D. Multipolar spindle pole coalescence is a major source of kinetochore mis-attachment and chromosome mis-segregation in cancer cells. PLoS ONE. 2009; 4(8). https://doi.org/10.1371/journal.pone.0006564 PMID: 19668340
14. Ganem NJ, Godinho SA, Pellman D. A mechanism linking extra centrosomes to chromosomal instability. Nature. 2009; 460(7252):278–282. https://doi.org/10.1038/nature08136
15. Fava LL, Schuler F, Sladky V, Haschka MD, Soratroi C, Eiterer L, et al. The PIDDosome activates p53 in response to supernumerary centrosomes. Genes and Development. 2017; 31(1):34–45. https://doi.org/10.1101/gad.289728.116 PMID: 28130345
16. Cosenza MR, Krämer A. Centrosome amplification, chromosomal instability and cancer: mechanistic, clinical and therapeutic issues. Chromosome Research. 2016; 24(1):105–126. https://doi.org/10.1007/s10577-015-9505-5
17. Lambrus BG, Uetake Y, Clutario KM, Daggubati V, Snyder M, Sluder G, et al. P53 protects against genome instability following centriole duplication failure. Journal of Cell Biology. 2015; 210(1):63–77. https://doi.org/10.1083/jcb.201502089 PMID: 26150389
18. Wong YL, Anzola JV, Davis RL, Yoon M, Motamedi A, Kroll A, et al. Reversible centriole depletion with an inhibitor of Polo-like kinase 4. Science. 2015; 348(6239):1155–1160. https://doi.org/10.1126/science.aaa5111 PMID: 25931445
19. Baudoin NC, Nicholson JM, Soto K, Martin O, Chen J, Cimini D. Asymmetric clustering of centrosomes defines the early evolution of tetraploid cells. eLife. 2020; 9. https://doi.org/10.7554/eLife.54565
20. Galofré C, Asensio E, Ubach M, Torres IM, Quintanilla I, Castells A, et al. Centrosome reduction in newly-generated tetraploid cancer cells obtained by separase depletion. Scientific Reports. 2020; 10 (1). https://doi.org/10.1038/s41598-020-65975-1 PMID: 32499568
21. Hill W. The Mathematical Theory of Selection, Recombination and Mutation. R. Burger. vol. 79. Chichester: Wiley; 2002.
22. Bettencourt-Dias M, Rodrigues-Martins A, Carpenter L, Riparbelli M, Lehmann L, Gatt MK, et al. SAK/PLK4 is required for centriole duplication and flagella development. Current Biology. 2005; 15 (24):2199–2207. https://doi.org/10.1016/j.cub.2005.11.042 PMID: 16326102
23. Lopes CAM, Jana SC, Cunha-Ferreira I, Zitouni S, Bento I, Duarte P, et al. PLK4 trans-Autoactivation Controls Centriole Biogenesis in Space. Developmental Cell. 2015; 35(2):222–235. https://doi.org/10.1016/j.devcel.2015.09.020 PMID: 26481051
24. Arquint C, Sonnen KF, Stierhof YD, Nigg EA. Cell-cycle-regulated expression of STIL controls centriole number in human cells. Journal of Cell Science. 2012; 125(5):1342–1352. https://doi.org/10.1242/jcs.099887
25. Strnad P, Leidel S, Vinogradova T, Euteneuer U, Khodjakov A, Gönczy P. Regulated HsSAS-6 Levels Ensure Formation of a Single Procentriole per Centriole during the Centrosome Duplication Cycle. Developmental Cell. 2007; 13(2):203–213. https://doi.org/10.1016/j.devcel.2007.07.004
26. Lens SMA, Medema RH. Cytokinesis defects and cancer. Nature Reviews Cancer. 2019; 19(1):32–45. https://doi.org/10.1038/s41568-018-0084-6

27. Moran PAP. Global stability of genetic systems governed by mutation and selection. II. Mathematical Proceedings of the Cambridge Philosophical Society. 1977; 81(3):435–441.

28. Eigen M. Selforganization of matter and the evolution of biological macromolecules. Die Naturwissenschaften. 1971; 58(10):465–523. https://doi.org/10.1007/BF00623322

29. Thompson CJ, McBride JL. On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules. Mathematical Biosciences. 1974; 21(1-2):127–142. https://doi.org/10.1016/0025-5564(74)90110-2

30. Eigen M, McCaskill J, Schuster P. Molecular quasi-species. Journal of Physical Chemistry 1988; 92, 24, 6881–6891. https://doi.org/10.1021/j100335a010

31. Sala R, Farrell K, Stearns T. Growth disadvantage associated with centrosome amplification drives population-level centriole number homeostasis. Molecular Biology of the Cell. 2020; p. mbc.E19–04–0195. https://doi.org/10.1091/mbc.E19-04-0195

32. Mukherji S, O'Shea EK. Mechanisms of organelle biogenesis govern stochastic fluctuations in organelle abundance. eLife. 2014; 2014(3). https://doi.org/10.7554/eLife.02678 PMID: 24916159

33. Wilke CO. Quasispecies theory in the context of population genetics. BMC Evolutionary Biology. 2005; 5. https://doi.org/10.1186/1471-2148-5-44

34. Lončarek J, Hergert P, Khodjakov A. Centriole reduplication during prolonged interphase requires procentriole maturation governed by plk1. Current Biology. 2010; 20(14):1277–1282. https://doi.org/10.1016/j.cub.2010.05.050

35. Serçin Ö, Larsimont JC, Karambelas AE, Marthiens V, Moers V, Boeckx B, et al. Transient PLK4 overexpression accelerates tumorigenesis in p53-deficient epidermis. Nature Cell Biology. 2016; 18 (1):100–110. https://doi.org/10.1038/ncb3270 PMID: 26595384

36. Levine MS, Bakker B, Boeckx B, Moyett J, Lu J, Vitre B, et al. Centrosome Amplification Is Sufficient to Promote Spontaneous Tumorigenesis in Mammals. Developmental Cell. 2017; 40(3):313–322.e5. https://doi.org/10.1016/j.devcel.2016.12.022 PMID: 28132847

37. Vitrea B, Holland AJ, Kulukian A, Shoshani O, Hirai M, Wanga Y, et al. Chronic centrosome amplification without tumorigenesis. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112(46):E6321–E6330. https://doi.org/10.1073/pnas.1519388112

38. Morretton JP, Herbette A, Cosson C, Mboup B, Latouche A, Gestraud P, et al. Centrosome amplification favours survival and impairs ovarian cancer progression. bioRxiv. 2019; p. 623983.

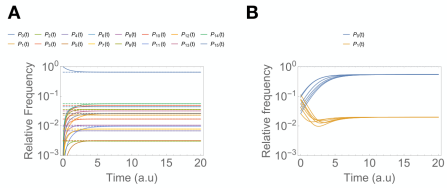
39. Kwon M, Godinho SA, Chandhok NS, Ganem NJ, Azioune A, Thery M, et al. Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes. Genes and Development. 2008; 22 (16):2189–2203. https://doi.org/10.1101/gad.1700908 PMID: 18662975

40. Godinho SA, Picone R, Burute M, Dagher R, Su Y, Leung CT, et al. Oncogene-like induction of cellular invasion from centrosome amplification. Nature. 2014; 510(7503):167–171. https://doi.org/10.1038/nature13277 PMID: 24739973

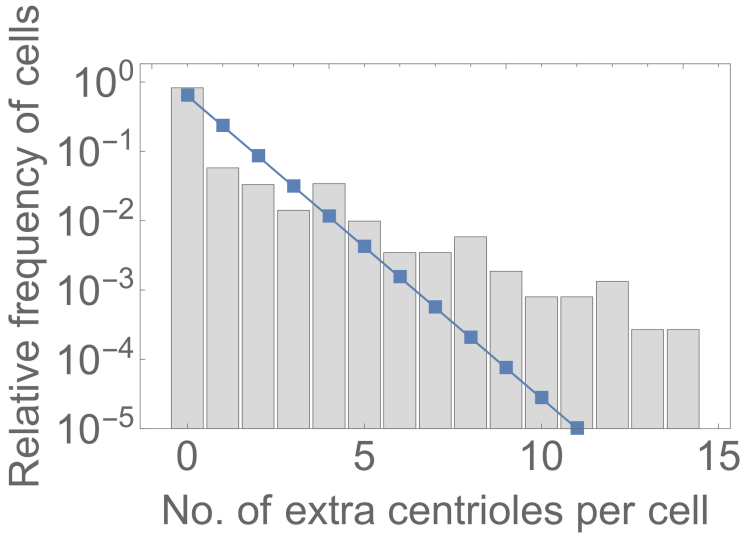
41. Arnandis T, Monteiro P, Adams SD, Bridgeman VL, Rajeeve V, Gadaleta E, et al. Oxidative Stress in Cells with Extra Centrosomes Drives Non-Cell-Autonomous Invasion. Developmental Cell. 2018; 47 (4):409–424.e9. https://doi.org/10.1016/j.devcel.2018.10.026 PMID: 30458137

42. LoMastro GM, Holland AJ. The Emerging Link between Centrosome Aberrations and Metastasis. Developmental Cell. 2019; 49(3):325–331. https://doi.org/10.1016/j.devcel.2019.04.002

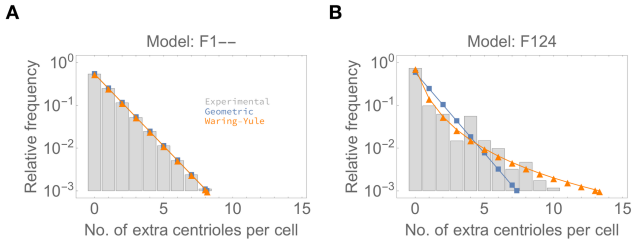
43. Denu RA, Zasadil LM, Kanugh C, Laffin J, Weaver BA, Burkard ME. Centrosome amplification induces high grade features and is prognostic of worse outcomes in breast cancer. BMC Cancer. 2016; 16(1). https://doi.org/10.1186/s12885-016-2083-x PMID: 26832928



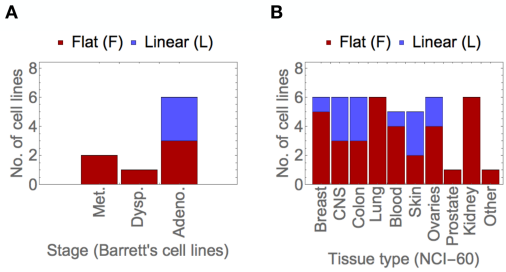
S1 Fig.: The solutions of the system of ordinary differential equations converge to the expected equilibrium value



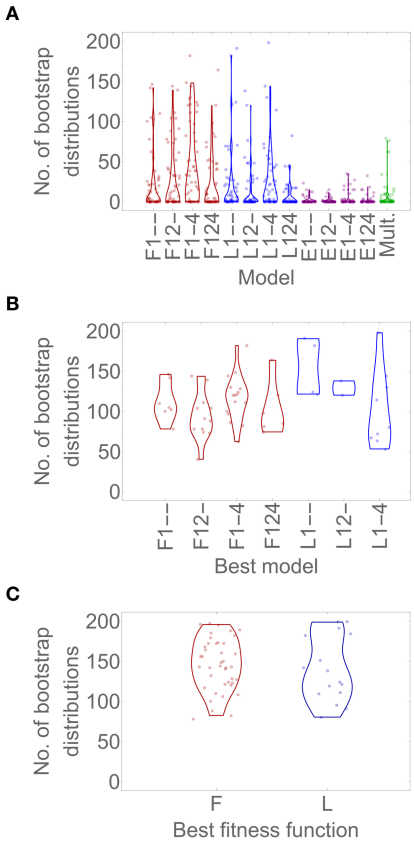
S2 Fig. Cells with high number of centrioles occur frequently across all data sets.



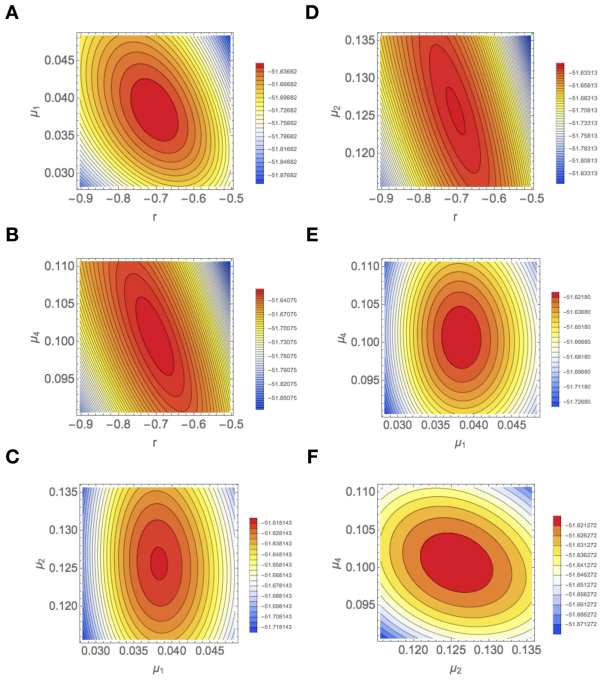
S3 Fig. Models in the candidate set can produce both geometric and heavy-tailed distributions.



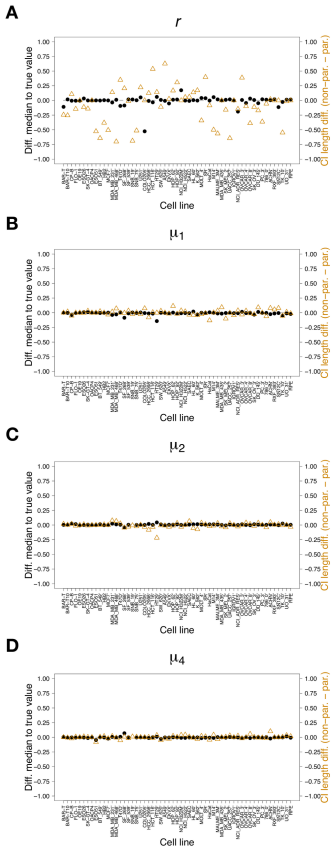
S4 Fig. Best fitness functions per tissue type.



S5 Fig. Bootstrap support of the candidate models.



S6 Fig. Model parameters are not correlated and the likelihood value shows little sensitivity to r.



S7 Fig. Parameter estimation is accurate and estimation errors are likely due to small sample sizes.

Chapter 3

A cell cycle-structured model of centriole number dynamics

Marco António Dias Louro[1,2,3], Matilde Palmeira[4], Mónica Bettencourt-Dias[1], Claudia Bank[2,3]

1. Instituto Gulbenkian de Ciência, Oeiras, Portugal

2. Institut für Ökologie und Evolution, Universität Bern, Bern, Switzerland

3. Swiss Institute of Bioinformatics, Lausanne, Switzerland

4. Departamento de Física, Universidade de Coimbra, Coimbra

Author contributions

This chapter is a preliminary draft for which we are conducting some additional analyses. We expect to submit it soon. The contributions of each author to the work were as follows: Marco António Dias Louro (the candidate), Mónica Bettencourt-Dias, and Claudia Bank conceptualised the study. Marco António Dias Louro developed the individual-based stochastic implementation of the model and the code for numeric simulations, generated the data and plots for figures 2, 4, and 6 and produced all the figures, and wrote the draft. Matilde Palmeira implemented the model as a system of ordinary differential equations and generated the data and produced all plots pertaining to figures 3 and 5, as well as all supplementary figures. Marco António Dias Louro, Matilde Palmeira, Mónica Bettencourt-Dias, and Claudia Bank edited the manuscript. Funding for this work was awarded to Marco António Dias Louro, Claudia Bank, and Mónica Bettencourt-Dias.

Abstract

Regulating the abundance of subcellular constituents is crucial for proper cell and organismal physiology. Centrioles are a special case in this regard because their copy numbers are exquisitely controlled. In typical proliferating cells, centrioles are duplicated once and only once per cell cycle and cells in which erroneous centriole biogenesis occurred are duly eliminated. This is essential to ensure a variety of processes which depend on centrioles, such as proper chromosome segregation. Deregulation of this copy number control has been widely associated with human disease. In particular, the occurrence of extra centrioles, or centrosome amplification, causes chromosome segregation errors and may lead to cancer development.

Several lines of evidence support the idea that centriole numbers in cancer cell populations are at an equilibrium emerging from the balance between centriole overproduction and selective elimination of cells with centrosome amplification. We previously showed that overproduction-selection balance can quantitatively explain a multitude of empirical distributions of centriole numbers per cell. However, centriole number dynamics depend on other processes, namely, centriole segregation at mitosis and cell cycle progression. Uncovering these dynamics is paramount for accurately scoring centriole number abnormalities and assessing their consequences for cellular physiology.

In this study, we developed a novel mathematical model of centriole number dynamics to assess how they depend on centriole segregation and the cell cycle. First, we confirmed that this model yields equilibrium distributions of centriole numbers per cell similar to the ones that were previously reported. Second, we quantified how model parameters affect centriole numbers at equilibrium and population fitness. As expected from overproduction-selection balance, the relative frequency of

cells with extra centrioles at equilibrium increases with the overproduction rate, whereas the fitness of the population decreases. Opposite trends were observed for increasing selective pressure against centrosome amplification. We observed that centriole segregation could increase fitness when it led to the asymmetric inheritance of extra centrioles, as experimentally observed, but only if selection acted in G1. Our model suggests that these asymmetric divisions are likelier for intermediate rates of centriole overproduction. Finally, we assessed how centriole numbers changed along the cell cycle. We observed that abnormal centriole numbers in interphase can be underscored when compared to mitosis, as commonly reported. However, our results suggest that the extent of this underscoring depended on the cell cycle stage at which cells with extra centrioles were eliminated. Taken together, our work allowed for a more thorough understanding of centriole number dynamics and their implications for human disease.

Keywords: centrosome amplification, modelling, centriole segregation, cell cycle

Introduction

How do cells regulate the copy number of their constituents? The abundance of organelles, such as mitochondria and plastids, depends on their rate of biogenesis, fusion, and fission. Such processes tend to lead to a steady-state distribution of organelle copy numbers in proliferating cell populations (Marshall, 2007, 2016; Mukherji and O'Shea, 2014). This process can be deregulated in diseases, such as cancer.

Centrioles represent an interesting case study in this regard. These intricate protein complexes are the core structures of the centrosome, the major centre for microtubule nucleation in many eukaryotic cell types. The canonical centrosome cycle proceeds as

follows: in G1, cells contain one centrosome with two centrioles, which undergo a single round of duplication during S-phase. After cell division, each daughter cell inherits a single centrosome with a pair of centrioles (Breslow and Holland, 2019; Brito et al., 2012; Gomes Pereira et al., 2021; Gönczy, 2015). Thus, the steady-state distribution of centriole numbers in healthy proliferating cells shows little variation - cells tend to possess two to four centrioles depending on the cell cycle stage. Notably, this is not true for cancer cell populations, in which centriole copy number tends to be abnormally high but extremely variable (Chan, 2011; Godinho and Pellman, 2014; Lopes et al., 2018; Marteil et al., 2018). Deviations to the wild-type number of centrioles generally impart a fitness cost for cells that bear them, triggering p53-dependent cell cycle arrest or multipolar cell divisions, which usually result in cell death (Burigotto et al., 2021; Coelho et al., 2015; Fava et al., 2017; Ganem et al., 2009; Holland et al., 2012; Kulukian et al., 2015; Kwon et al., 2008; Levine et al., 2017; Serçin et al., 2016; Tkach et al., 2022; Vitre et al., 2015; Wong et al., 2015).

Notwithstanding, even these abnormal distributions of centriole numbers per cell appear to be at equilibrium. Several studies reported that the initial distribution of centriole numbers per cell was recovered after transient perturbation. For instance, transient overexpression of an essential centriole biogenesis regulator, Plk4, treatment with cytokinesis-blocking drugs followed by their washout, or transient inhibition/degradation of Plk4, produced short-term centriole number change which rapidly revert to the initial state, as long as cell proliferation was not irreversibly halted (Baudoin et al., 2020; Galofré et al., 2020; Lambrus et al., 2015; Sala et al., 2020; Wong et al., 2015). These reports suggest that centriole copy number is maintained by a balance between centriole (over)production and negative selection against cells carrying abnormal centriole numbers. Here, we refer to this

phenomenon as overproduction-selection balance.

Previous work in our lab found that empirical distributions of centriole numbers per cell at mitosis can be explained by quantitative models of overproduction-selection balance. We inferred generally strong selection against extra centrioles and complex overproduction events in which multiple centrioles can be gained in a single-step. Our results indicated that any number of extra centrioles was equally deleterious, whereas intuitively one might expect that a higher number of extra centrioles equals lower fitness (Dias Louro et al., 2021). This raised the possibility that for quantifying centriole number dynamics it is sufficient to consider wild-type-like and abnormal cells.

However, our previous work made some simplifications. First, it ignored how extra centrioles were segregated after cell division. The data on this subject is scarce but seems to suggest that extra centrioles are segregated randomly (Sala et al., 2020; Wang et al., 2011). This can lead to both asymmetric and symmetric divisions, in which daughter cells can inherit, different or identical numbers of extra centrioles, respectively. Asymmetric divisions facilitate chromosome segregation errors and aneuploidy (Cosenza et al., 2017). In some cases, asymmetric divisions can result in one daughter cell inheriting *all* extra centrioles present in the mother cell, whereas the other will enter the new cell cycle with two centrioles. These cells tend to be more viable compared to cells which have inherited extra centrioles. Indeed, asymmetric divisions were found to improve population fitness following transient centrosome amplification caused by cytokinesis failure (Baudoin et al., 2020). Whereas asymmetric divisions can be advantageous for cells undergoing malignant transformation as they might facilitate the survival of aneuploid progeny, it is not known how often they occur under random centriole segregation and if their benefits hold under constant centriole overproduction.

Secondly, since the data set we previously analysed included

only mitotic cells, there was no need to take centriole number dynamics along the cell cycle into consideration. As previously mentioned, centriole biogenesis tends to occur at a specific point in the cell cycle. Conversely, the selective mechanisms responsible for culling cells carrying abnormal centriole numbers can act at different time points along the cell cycle. Some mechanisms, such as p53-dependent cell cycle arrest and cell death followed by multipolar divisions, tend to occur in interphase. Others occur during mitosis, such as spindle assembly checkpoint-dependent delay of anaphase (Ganem et al., 2009; Sala et al., 2020). Thus, centriole overproduction can affect the rate of cell cycle progression. Conversely, changes to the distribution of cell cycle phases may affect the distribution of centriole numbers per cell in the population. This can lead to a complex interaction between centriole overproduction and cell cycle dynamics which lacks quantitative understanding.

Here, we extended our previous modelling approach to account for post-mitotic centriole inheritance and centriole number changes during the eukaryotic cell cycle. The basic structure of the model was rooted in the traditional representation of the eukaryotic cell cycle, consisting of four phases: G1 (gap-1), S, G2 (gap-2), and M (mitosis). Centriole duplication and random overproduction were assumed to occur at the transition between G1 and S. After mitosis, each daughter cell inherited a random number of extra centrioles. First, we confirmed that the model produced heavy-tailed distributions as previously observed in a broad data set. Second, we explored the consequences of biased centriole segregation, yielding asymmetric divisions exclusively, and unbiased, or random centriole segregation under constant centriole overproduction. Our results indicated that biased centriole segregation marginally improved the growth rate of simulated populations only for extreme centriole overproduction rates. In addition, our model predicts that the probability of extremely

asymmetric divisions is maximised for intermediate centriole overproduction rates. Third, we investigated the interplay between centriole overproduction and the cell cycle. We showed that the timing of selection may lead to systematic underscoring of centriole number abnormalities in interphase cells, which requires knowledge of the underlying cell cycle-associated processes to correct. Finally, we discuss the implications of this work for the role of centrosome amplification in tumourigenesis and for the use of centriole numbers as biomarkers for cancer.

A model of centriole number dynamics along the cell cycle

We developed a new model for describing centriole number dynamics along the cell cycle to extend our previous formalism (Figure 3.1). Crucially, this model explicitly differentiates between centriole biogenesis and post-mitotic inheritance. Our subject of focus is a population of proliferating human cells growing in an unlimited environment. Each cell is fully characterised by the number of extra centrioles $i = 0, 1, 2...$ it harbours and its current cell cycle stage at time t.

Based on classical mathematical models models of the eukaryotic cell cycle (Ritter et al., 1994; Weber et al., 2014; Yanagisawa et al., 1985), we posit that cells can progress through G1, S, G2, and M, at which point they divide, yielding two daughter cells in G1. The mean residence time of each cell cycle stage is given by τ_{G1}, τ_S, τ_{G2}, and τ_M. It can be shown that such a model eventually yields a population undergoing exponential growth, with the relative frequency of cells in each cell cycle stage being at an equilibrium, for a non-zero initial population size. We assume perfect centriole duplication, which occurs instantaneously as cells progressed from G1 to S-phase. After cell division, each daughter cell inherits at least two centrioles. Perfect duplication

and segregation of wild-type centriole numbers ensures that cells contain at least two or four centrioles depending on the stage of the cell cycle.

We consider that centriole overproduction occurs via two concurrent "pathways". First, all extra centrioles present in G1 undergo duplication as the cell transitions into S-phase. Second, we assume random centriole overproduction according to a Poisson distribution with mean/rate β - meaning that a cell produces β extra centrioles, on average. Heretofore, β will be referred to as the centriole overproduction rate. While β can take any non-negative value, we consider a range of $[0, 1]$. In the lower bound (β=0), cells do not produce any extra centrioles. In the upper bound (β=1), cells produce one extra centriole on average. For practical purposes, we assume that cells could attain at most i_{max} centrioles, which was necessary for analysing the models numerically.

Concerning post-mitotic centriole inheritance, we assume that each extra centriole was segregated randomly, with probability α. If $\alpha = 0.5$, an extra centriole can be inherited by either daughter cell with equal probability. If $\alpha = 0$ or $\alpha = 1$, said extra centriole would be always inherited by the same daughter cell. This allowed us to study model dynamics under unbiased ("random", $\alpha = 0.5$) or biased centriole segregation ($\alpha = 0$).

Finally, we assume that negative selection acted on cells carrying extra centrioles, such that they experience a cell cycle delay proportional to w_{G1}, w_S, w_{G2}, and/or w_M, depending on the stage at which selection acts. These parameters are correctly interpreted as the fraction of cells that "arrested" in the respective cell cycle stage, in a given time-window. Alternatively, they can be interpreted as a proportional increase in the mean residence time at each cell cycle stage, e.g. as w_{G1} increases, cells with extra centrioles take longer to progress through G1. Based on our previous work, we made the simplifying assumption that selection did

not change with the number of extra centrioles. We focus on two extreme selection regimes (Dias Louro et al., 2021). First, we refer to cell cycle stage-specific selection when only one of w_{G1}, w_S, w_{G2} or w_M has a non-zero value, i.e. cells are delayed/arrested only in that stage. Second, we refer to cell cycle-unspecific selection when. $w = w_{G1} = w_S = w_{G2} = w_M$), i.e. selection is equally strong in each cell cycle stage. Note that if the value of these parameters is too low relative to the centriole overproduction rate β, centriole number will increase indefinitely (or up to the maximum number of extra centrioles per cell, i_{max}. Thus, we varied w_{G1}, w_S, w_{G2} and/or w_M in the range $[0.4, 0.9]$, which is sufficient to ensure an equilibrium state.

We describe the temporal dynamics of the relative frequencies of each sub-population in an infinitely-sized population, containing i extra centrioles, which we solved numerically. Additional biological and mathematical information on this model can be found in the 3 and 3 sections.

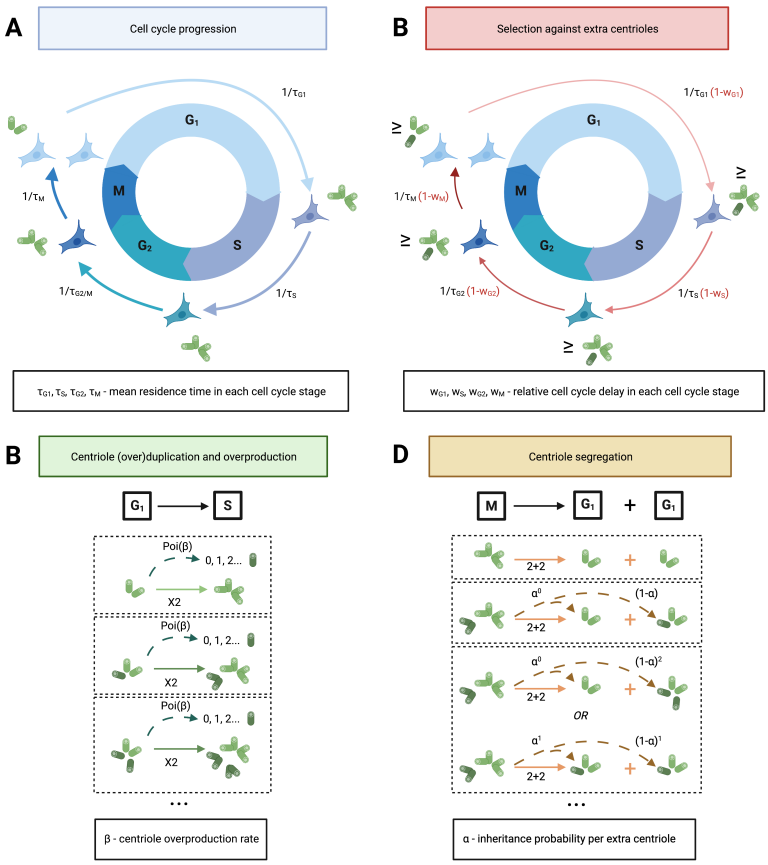


FIGURE 3.1: **Model representation.** A - The human cell cycle was assumed to comprise four distinct stages - G1, S, G2, and M (mitosis). Cells progress through each cell cycle stage at a rate $1/\tau_{G1}$, $1/\tau_S$, $1/\tau_{G2}$, and $1/\tau_M$. When exiting mitosis, two daughter cells are produced. B - Selection against extra centrioles is envisioned as triggering cell cycle delay/arrest. In practice, the rate of cell progression in cells with extra centrioles is proportional to fraction the fraction of cells with wild-type centriole numbers, equal to $1 - w_{G1}$, $1 - w_S$, $1 - w_{G2}$, and $1 - w_M$. C - Centriole duplication is depicted as an instantaneous and simultaneous process at the G1-S transition. We posit that all preexisting centrioles duplicated with perfect fidelity. We include random centriole overproduction as a Poisson process with mean value β. D - Upon cell division, we assume that each daughter cell inherits at least two centrioles. Extra centrioles, on the other hand, segregate randomly as a Bernoulli trial, with probability α - i.e. they are either inherited by one daughter cell or the other.

Results

The model can generate heady-tailed distributions of centriole numbers per cell at mitosis

A peculiar feature observed in previous experimental studies was that the distribution of centriole numbers per cell was heavy-tailed. Such distributions could be generated by multi-step overproduction events (Dias Louro et al., 2021). Heavy-tailed distributions are probability distributions in which extreme values are overrepresented compared to an exponential distribution, for continuous data, or to a geometric distribution, for discrete data. In the context of this study, a distribution is heavy-tailed if cells with high numbers of centrioles are more frequent than what would be expected from a geometric distribution.

To test if this model could generate heavy-tailed distributions, we computed equilibrium distributions of centriole numbers per cell at mitosis and compared them to geometric distributions (Fig. 3.2). We assumed that centriole inheritance to be unbiased ($\alpha = 0.5$) and strong, uniform, negative selection against extra centrioles in all cell cycle stages ($w_{G1} = w_S = w_{G2} = w_M$). The results showed that the model can, indeed, generate heavy-tailed distributions (Fig. 3.2B and C). Moreover, these distributions display a characteristic pattern in which even numbers of extra centrioles are likelier than the nearest odd numbers. This pattern was previously observed in empirical data. Increasing the centriole overproduction rate β produced a right-shift in the equilibrium distributions, i.e. centriole number increased, on average, but they remained heavy-tailed.

Since we previously showed that heavy-tailed distributions could be generated by multi-step centriole overproduction events (Dias Louro et al., 2021), we asked if duplication of all preexisting centrioles was sufficient to elicit the same behaviour in this model. To test this, we simplified the model such that only a single pair of

centrioles could duplicate (Fig. 3.2D). Thus, extra centrioles arose exclusively from Poisson-distributed overproduction events. In this case, the model yielded distributions with light-tailed distributions, i.e. cells with extremely high centriole numbers were rarer than in a geometric distribution (Fig. 3.2D and E)). As before, higher values of the centriole overproduction rate, β, the distribution was shifted to the right without affecting its tailedness. We concluded that this model can generate heavy-tailed equilibrium distributions of centriole numbers per cell when all preexisting centrioles are allowed to duplicate, recapitulating previous empirical observations.

Fitness and extra centriole abundance showed unexpected dependencies on model parameters

Next, we proceeded by evaluating the sensitivity of model behaviour to three key parameters: the centriole overproduction rate, β, the unspecific cell cycle delay, w, and the inheritance probability, α. We were especially interested in tracking two response variables: the relative equilibrium frequency of extra centrioles at mitosis, which is in indicate of the extent of centrosome amplification, and the exponential growth rate of the population, as a proxy for fitness.

As expected, the relative frequency of mitotic cells at equilibrium increased with the centriole overproduction rate β and tended towards 1 (i.e. all mitotic cells containing extra centrioles, Fig. 3.3B). Concomitantly, the mean number of extra centrioles in mitotic cells also increased at an faster rate - a consequence of the heavy-tailedness of the distributions (Fig. S1A). However, increasing the strength of selection (higher w) had little effect in the relative abundance of extra centrioles in the population (Fig. 3.3B), but greatly affected their mean number (Fig. S1B). In turn, the exponential growth rate decreased linearly with both the centriole overproduction rate β and relative cell cycle delay, w (Fig.

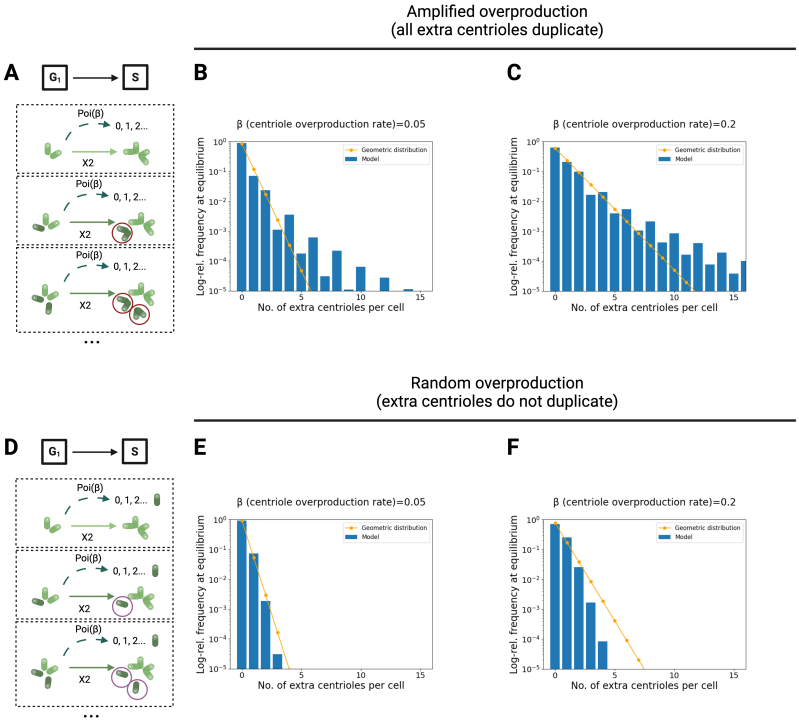


FIGURE 3.2: **Duplication of all extra centrioles was sufficient to generate heavy-tailed distributions of centriole numbers per cell at mitosis.** Depiction of centriole biogenesis with duplication of all extra centrioles (A) and only wild-type centriole numbers (D). Simulated distributions of centriole numbers per cell at mitosis (blue bars) compared to the best fitting geometric distribution (orange points/line) assuming low (B, E) and high rates of overproduction (C, F). The default parameter values were $w = 0.9$ and $\alpha = 0.5$.

3.3A). In summary, the relative frequency of extra centrioles was especially sensitive to the rate of overproduction but selection was inefficient in purging cells with extra centrioles from the population. Nevertheless, selection against extra centrioles considerably limited population growth and reduced mean centriole numbers in mitotic cells at equilibrium.

As mentioned above, asymmetric divisions can improve population fitness by allowing one daughter cell to inherit wild-type centriole numbers. This phenomenon observed in populations

of cells treated with a cytokinesis-blocking drug, which led to a transiently doubling of centriole numbers in the affected cells (Baudoin et al., 2020). However, in most experimental systems, cells are subject to different degrees of centriole overproduction (Lopes et al., 2015; Lopes et al., 2018; Marteil et al., 2018). Therefore, it is pertinent to ask if asymmetric divisions can confer a selective advantage to cells in this context. In our model, the probability of asymmetric divisions for cell survival can be modulated by varying the inheritance probability, α. When $\alpha = 0.5$, centriole segregation is entirely random or unbiased, such that each extra centriole can be inherited by either daughter cell with equal probability. In contrast, $\alpha = 0$ or $\alpha = 1$ represents completely biased segregation, in which a single daughter cell inherits all extra centrioles. In other words, all cell divisions are completely asymmetrical. Values in between 0 and 0.5 (or between 0.5 and 1) can represent as intermediate degrees of bias.

Our results showed that neither the relative frequency of mitotic cells with extra centrioles at equilibrium nor the exponential growth rate were nearly insensitive to α for low centriole overproduction rates (Fig. 3.3E,F). For high centriole overproduction rates, the relative frequency of mitotic cells with extra centrioles decreased slightly and the exponential growth rate suffered a modest increased as centriole segregation became more biased, i.e. when asymmetric divisions were favoured. In contrast, mean number of extra centrioles at equilibrium decreased monotonically as α increased (Fig. S1C). These results seem to suggest that asymmetric divisions are not an effective coping mechanism for centrosome amplification under constant centriole overproduction, except perhaps for extreme centriole overproduction rates. To understand this, note that we assumed that selection acted uniformly all throughout the cell cycle. Thus, these results can be due to the fact that asymmetric divisions can only "relieve" G1 cells from the burden of extra centrioles, which might still

suffer a fitness penalty if they undergo overproduction in the subsequent S-phase.

To better understand how the timing of selection affected model behaviour, we asked how the exponential growth rate and the equilibrium relative frequency of extra centrioles at mitosis changed with the parameter values when selection acted specifically on G1 (Fig. S2, S (Fig. S4), G2 (Fig. S6), and M (Fig. S8), respectively). We observed qualitatively similar trends when we varied the centriole overproduction rate and each of the stage-specific relative cell cycle delays - higher overproduction rates increased the relative equilibrium frequency of mitotic cells with extra centrioles and decreased the exponential growth rate whereas stronger selection reduced both. Mean centriole number also showed similar trends compared to the unspecific selection case (Fig. S3, Fig. S5, Fig. S7, S9). However, the results differed concerning the inheritance probability. In particular, when selection acted in G1, we observed that the exponential growth rate increased when centriole segregation was more biased and the overproduction rate was high (Fig. S2E). In this case, one daughter cell was guaranteed to escape the full effects of selection, regardless if it acquired extra centrioles in the following S-phase. Thus, asymmetric divisions can be beneficial for the population if selection against centrosome amplification acts specifically in G1.

In summary, the centriole overproduction rate was the key parameter for determining the equilibrium frequencies of mitotic centriole numbers, and their mean. Selection could effectively decrease the rate of proliferation of cells carrying extra centrioles but it could not eliminate them - only decrease the mean number of extra centrioles. Strikingly, the inheritance probability, ranging from entirely random/unbiased to biased centriole segregation, had minimal impact on the relative frequency of extra centrioles, mean centriole number, or fitness if selection was not specific to any cell cycle stage. However, biased centriole segregation,

yielding mostly asymmetric divisions could be a viable adaptive strategy if selection acted specifically in G1. This is likely to be the case in a variety of conditions, namely, in the presence of active p53.

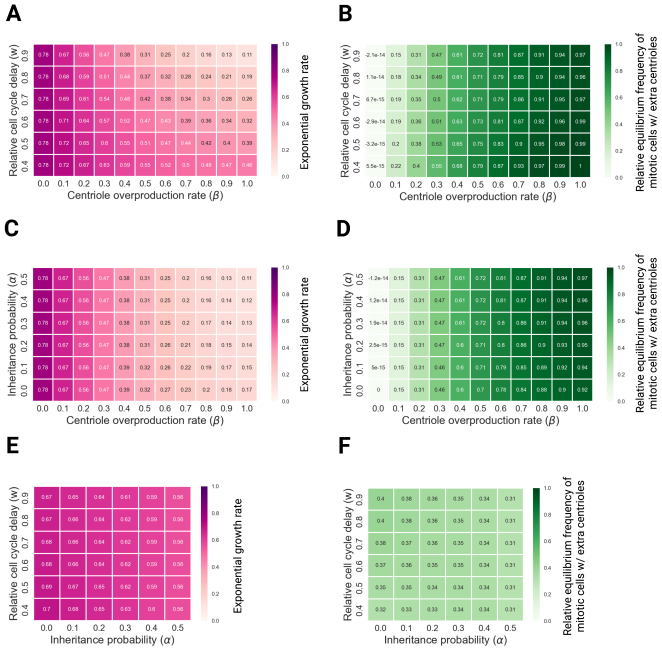


FIGURE 3.3: **The exponential growth rate and degree of centriole number anomalies depended chiefly on the overproduction rate.** Biparametric heatmaps of the exponential growth rate (A, C, E) and relative frequency of mitotic cells with extra centrioles at equilibrium (B, D, F) as a function of each pair of parameters. A, B - Centriole overproduction rate versus unspecific relative cell cycle delay; C, D - Centriole overproduction rate versus inheritance probability; E, F - Inheritance probability versus unspecific relative cell cycle delay. The default parameter values were $\beta = 0.2$, $w = 0.9$, $\alpha = 0.5$.

Asymmetric divisions were more frequent for intermediate overproduction rates

The results above suggested that asymmetric divisions can be advantageous in a specific selective regime. In addition, there is

data indicating that asymmetric divisions increase the frequency of chromosome missegregation events, leading to aneuploid offspring (Cosenza et al., 2017. Therefore, asymmetric divisions can potentially contribute to cancer development. Since centriole segregation tends to occur randomly, we asked how likely it is for a cell to undergo an asymmetric division in this setting. Note that we refer to asymmetric divisions as a bipolar mitosis in which a cell with centrosome amplification yields one daughter cell with wild-type centriole numbers and another containing all preexisting extra centrioles. Per this definition, a cell containing wild-type centriole numbers cannot undergo an asymmetric division in this model.

We started by calculating the probability that a mother cell containing i extra centrioles at mitosis generates a daughter cell with wild-type centriole numbers. Under random/unbiased centriole segregation, the probability of undergoing an asymmetric division decreases exponentially as the number of extra centrioles in the mother cell increases (Fig. 3.4B). This can be intuitively understood because the more extra centrioles there are, the less likely it is that all of them are randomly segregated into *the same* daughter cell. Such a phenomenon has been experimentally observed (Sala et al., 2020). Conversely, if centriole segregation is completely biased, the probability of undergoing an asymmetric division is equal to 1, regardless of the number of extra centrioles (Fig. 3.4B).

Next, we calculated the relative frequency of wild-type-like daughter cells stemming from an asymmetric division in a population of cells at equilibrium. When the centriole overproduction rate, β, increased the relative frequency of cells containing wild-type centriole was maximised for some intermediate value of β. This trend can be explained by the fact that for lower values of the overproduction rate, there were few cells with extra centrioles in the population. Therefore, as the overproduction rate increased,

so too did the probability of undergoing an asymmetric division. To understand the decrease, it should be noted that asymmetric divisions were likelier for cells with low numbers of extra centrioles (Fig. 3.4B). Thus, after a certain threshold, increasing the overproduction rate led to cells containing, on average, a high number of extra centrioles, which were less likely to undergo an asymmetric division. Coherently, this threshold shifted to the right as the strength of unspecific selection, w increased (Fig. 3.4C). If selection was weaker, extra centrioles accumulated more easily in the population, and thus, cells with low numbers of extra centrioles were more common for low overproduction rates. If selection was stronger, then higher overproduction rates were necessary to increase the relative frequency of cells with low numbers of extra centrioles.

For comparison, when centriole segregation and inheritance were biased ($\alpha = 0$, Fig. 3.4D), the relative frequency of wild-type daughter emerging from asymmetric divisions increased with the overproduction rate (Fig. 3.4F). Trivially, if all cells in the population contained extra centrioles, all cell divisions would yield one daughter with wild-type centriole numbers and another one which inherited all extra centrioles (Fig. 3.4E). As previously observed, selection counteracted centriole overproduction by reducing the equilibrium frequency of cells with extra centrioles (Fig. 3.4F).

The cell cycle profile changed even if the timing of selection was unspecific

Centrosome amplification can trigger cell cycle delay or arrest in different stages of the cell cycle (Burigotto et al., 2021; Fava et al., 2017; Ganem et al., 2009; Holland et al., 2012). This process can lead to altered cell cycle profiles in proliferating cell populations. The extent of these alterations should be proportional to the frequency of cells with centrosome amplification. To explore this

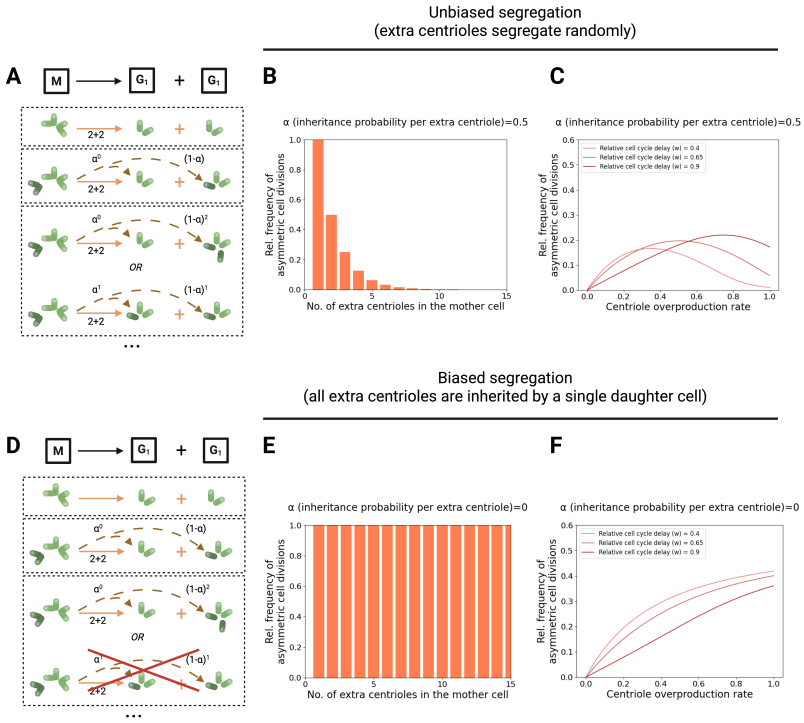


FIGURE 3.4: **Asymmetric divisions were maximised for intermediate rates of centriole overproduction.** A - Cell division outcomes under unbiased (random) centriole segregation. Note that cells with two extra centrioles can generate three types of daughter cells - containing either zero, one, or two extra centrioles. B - Probability of asymmetric divisions under unbiased centriole segregation as a function of the number of centrioles in the mother cell. C-Relative frequency of asymmetric divisions as a function of the overproduction rate, assuming weak (light red), moderate (red), and strong (dark red) unspecific relative cell cycle delays and unbiased centriole segregation. D - Cell division outcomes under biased centriole segregation. Note that the mother cell can only yield a daughter cell with zero extra centrioles and another one containing all extra centrioles. D - Probability of asymmetric divisions under biased centriole segregation as a function of the number of centrioles in the mother cell. E - Relative frequency of asymmetric divisions as a function of the overproduction rate, assuming weak (light red), moderate (red), and strong (dark red) unspecific relative cell cycle delays and biased centriole segregation.

phenomenon, we asked how the underlying cell cycle dynamics changed as a function of model parameters. We computed the

equilibrium cell cycle profiles assuming unspecific selection and selection acting specifically on G1, S, G2, and M and studied how they varied with model parameters (Fig. 3.5).

When selection acted non-specifically along the cell cycle, the relative frequencies at equilibrium of G1- and S-phase cells exhibited antagonistic behaviours as the centriole overproduction rate, β, increased (Fig. 3.5A). Initially, G1-phase cells became rarer with increasing overproduction rates, and S-phase cells rose in abundance. At a critical value of the overproduction rate, the trend reversed - cells in G1 increased in frequency and cells in S decreased. This behaviour can be understood by considering extreme cases - if the overproduction rate is zero, all cells would have wild-type centriole numbers. If it is high enough, nearly all cells would have extra centrioles. When selection acts equally across all stages, the equilibrium cell cycle profiles will be identical in these two cases. For intermediate overproduction rates, some cells would have extra centrioles whereas others would not. Cells with extra centrioles are instantly delayed in S-phase as soon as overproduction occurs, leading to an accumulation of cells in that stage. Since cell cycle delay would lead to fewer cell divisions, G1 cells are depleted. We also observed accumulation of S-phase cells with increasing strength of selection, i.e. stronger cell cycle delay (Fig. 3.5B). In contrast, changing the inheritance probability, ranging from unbiased (random) segregation to biased segregation had little effect in the equilibrium cell cycle profile (Fig. 3.5C).

When selection against extra centrioles acted specifically at a certain cell cycle stage, we observed that the relative frequency of cells in the stage at which selection operated increased as a function of the overproduction rate and the strength of selection, and when centriole segregation was more biased (Fig. 3.5D-O). whereas the others decreased. For instance, if selection acted only in G1, cells tended to accumulate in that stage.

In conclusion, increasing the rate of centriole overproduction or the strength of selection led to shifts in the equilibrium cell cycle profile. More specifically, the population was overrepresented in the cell cycle stage at which selection acted, or at the stage of centriole biogenesis in the case of unspecific selection. Concerning the latter, we observed an excess of S-phase cells only for intermediate overproduction rates. In contrast, the equilibrium cell cycle profile was identical when comparing populations experiencing no centriole overproduction or extremely high centriole overproduction rates. Interestingly, changing the inheritance probability affected equilibrium cell cycle profile only when selection against extra centrioles was stage-specific.

The timing of selection can lead to unpredictable underestimation of centriole number anomalies

We have established that the rate of centriole overproduction, the strength of selection, and the timing of selection could affect both the equilibrium distributions of centriole numbers per cell in mitosis and the equilibrium cell cycle profile. What could be the consequences for an interphase cell population? We were motivated to answer this question because centriole number abnormalities have often been scored in non-mitotic cells, especially in tumour biopsies (Bettencourt-Dias et al., 2011). An interphase cell population consists of a mix of cells in G1, S, and G2 (i.e. non-mitotic), whose precise cell cycle stage is often not known. Since centriole biogenesis took place at the onset of S-phase in our model, if the distribution of cell cycle stages changes, then the distribution of extra centrioles per cell should also change.

In the absence of additional markers specific to a given cell cycle stage, one can only assure that a cell containing more than four centrioles is abnormal (Bettencourt-Dias et al., 2011). This criterion leads to underscoring, since a G1 cell containing three or four centrioles is abnormal too. It can be readily seen in our

model that the relative frequency of G1 cells containing three or four centrioles increased with the overproduction rate (Fig. 3.5A-B). This suggests that the degree of underscoring depends on the degree of errors in centriole biogenesis.

To address this possibility systematically, we calculated the difference between equilibrium frequencies of cells with more than four centrioles at mitosis and interphase (Fig. 3.5). For unspecific selection, the underscoring of interphase cells with centriole number anomalies increased with the strength of selection and with moderate overproduction rates (Fig. 3.5C). For extreme overproduction rates, the difference between mitotic and interphase cells decreased because most interphase cells harboured more than four centrioles.

These results were qualitatively similar to the case when selection acted specifically in G1. When selection acted specifically in S-phase or G2, the difference between mitotic and interphase cells was greater when the overproduction rate was moderate and the relative cell cycle delay was weak to moderate but the effect was significantly smaller than in the other cases. Interestingly, if selection acted specifically in mitosis, the difference between mitotic and interphase cells was maximal for low overproduction rates and strong selection. This is due to cells with low numbers of extra centriole accumulating in mitosis, which remain undetected in the ensuing stages. In the whole range of parameter values that we studied, centriole number anomalies were systematically underscored in interphase. Taken together, we conclude that centriole number anomalies can be severely underscored in interphase. To correct this underscoring, it is necessary to know when selection against extra centrioles tends to occur.

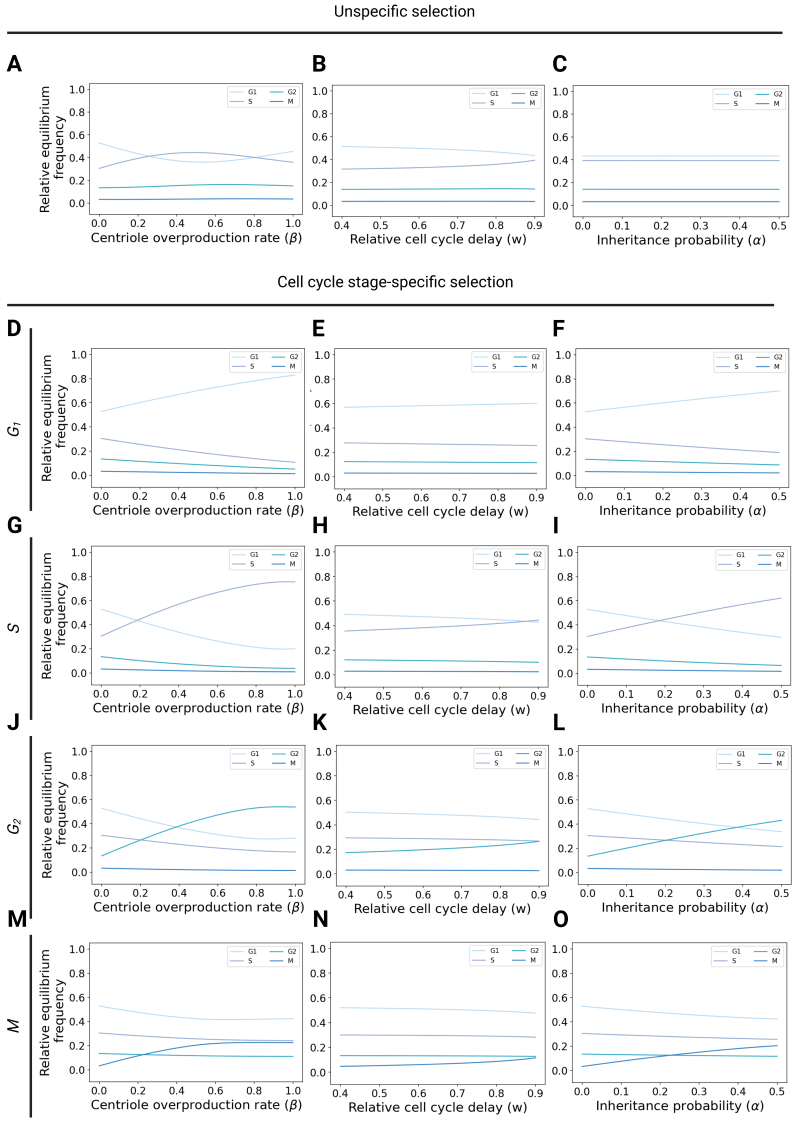
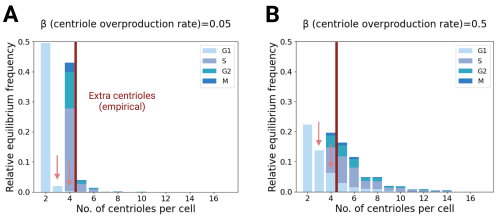
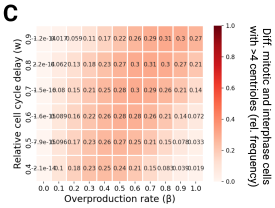


FIGURE 3.5: **The equilibrium cell cycle profiles changed depending on the timing of selection** Relative equilibrium frequencies of cells in G1 (light blue), S (purple), G2 (turquoise), and M (dark blue) as a function of the overproduction rate (A, D, G, J, M), relative cell cycle delay (B, E, H, K, N), and inheritance probability (C, F, I, L, O). Negative selection was assumed to affect all cell cycle stages equally (A-C) or acted exclusively in G1 (D-F), S (G-I), G2 (J-L), or M (M-O). The default parameter values were $\beta = 0.2$,$w = 0.9$ (A-C) $w_G1 = 0$ (G-O), $w_S = 0$ (D-F,J-O), $w_G2 = 0$ (D-I,M-O), $w_M = 0$ (D-L), and $\alpha = 0.5$.



Unspecific selection



Cell cycle stage-specific selection

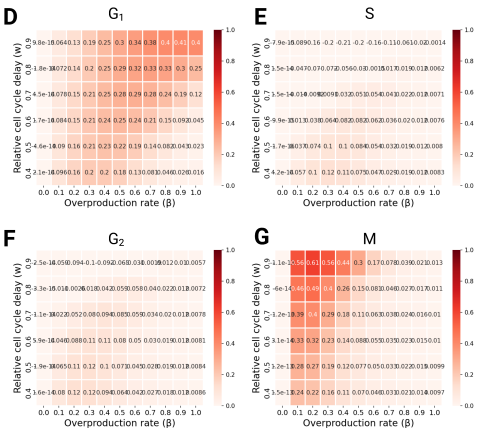


FIGURE 3.5 *(previous page)*: **The timing of selection can lead to unpredictable underestimation of centriole number anomalies.** A, B: Histograms of centriole numbers per cell in each cell cycle stage at equilibrium (G1 - light blue; S - blue; G2 - blue-green; M - dark blue). The red line indicates that abnormally high centriole numbers can only be reliably scored in interphase cells if those cells contain more than four centrioles. The red arrow indicates G1 cells with extra centrioles (3) that would not be scored as abnormal according to the empirical criterion. A - Centriole overproduction rate β = 0.05. B - Centriole overproduction rate β = 0.5. The default parameter values were $w = 0.9$, $\alpha = 0.5$. C-G - Biparametric heatmaps of the difference in relative frequency of mitotic cells with extra centrioles and interphase cells with extra centrioles as a function of the overproduction rate and relative cell cycle delay assuming unspecific selection (C, $w_{G1} = w_S = w_{G2} = w_M$), selection acting specifically in G1 (D), selection acting specifically in S (E), selection acting specifically in G2 (F), and selection acting specifically in M (G). The default parameter values were $\alpha = 0.5$, $w_{G1} = 0$ (D, E, F), $w_S = 0$ (C, E, F), $w_{G2} = 0$ (C, D, F), and $w_M = 0$ (C, D, E).

Discussion

Here, we developed a novel mathematical model for quantifying the effects of centriole biogenesis and centriole segregation along the cell cycle. Our main goal was to foster a quantitative understanding of how multiple factors, such as the rate of centriole overproduction and the timing of selection, could affect empirical distributions of centriole numbers per cell and cell cycle profiles at equilibrium. To this end, we sought to inspect biologically plausible scenarios based on assumptions with extensive empirical backing. Indeed, we confirmed that our model can recapitulate the general properties of observed distributions of centriole numbers per cell - a stable equilibrium state and heavy tailed distributions of centrioles numbers in mitotic cells.

Our main findings can be summarised as follows: (1) centriole number abnormalities were particularly sensitive to the rate of centriole overproduction, and less so to the strength of selection; (2) asymmetric segregation of extra centrioles after mitosis can improve fitness if selection against cells with centrosome amplification occurs in G1; (3) the probability of asymmetric divisions is maximised for intermediate centriole overproduction rates; (4) cell cycle profiles at equilibrium depended on when selection acted; (5) centriole number anomalies in interphase cells can be underscored to different extents depending on the timing of selection. In the following, we discuss the implications of the results for a quantitative understanding of centriole number dynamics and their consequences.

Limitations of this model

The modelling framework here presented addressed some important simplifications of our previous approach. Namely, this model fully encompassed centriole segregation and cell cycle dynamics. In addition, the focus on a population-level description

allowed us to rely on solid parameter estimates for the rates of cell cycle progression, centriole segregation, and selection against extra centrioles.

As discussed above, the centriole assembly pathway is considerably complex, even though to the best of our knowledge, it revolves around a small number of molecular species (Arquint et al., 2012; Bettencourt-Dias et al., 2005; Kleylein-Sohn et al., 2007; Strnad et al., 2007). Thus, summing up centriole biogenesis as random centriole overproduction together with duplication of all extra centrioles may appear somewhat artificial, as both processes are bound to be linked. However that may be, this approach can be duly justified by existing data Marshall, 2007; Sala et al., 2020). Random overproduction with each preexisting centriole enacting a positive feedback might be a more natural way to implement the same process, but additional data is necessary to verify such an assumption. Whichever the case, we expect that a positive feedback on centriole overproduction or explicitly assuming duplication of all extra centrioles and random centriole overproduction should yield similar results in most relevant conditions.

Finally, our model did not consider multipolar divisions or cytokinesis failure. Whereas the former might have circumstantial importance, since most multipolar divisions tend to be lethal, the latter might be more relevant. In fact, a significant amount of cancers experience whole-genome duplication, which might stem from cytokinesis failure (Lau and Poon, 2023). In the context of our modelling approach, cytokinesis failure could produce: (1) no changes in population size, i.e. a cell would simply transition from M to G1; (2) a precisely determined change in centriole numbers, i.e. a cell in M-phase containing i extra centrioles would give rise to a G1 cell with i extra centrioles. If cytokinesis failure is rare and confers no other selective advantage to cells, its importance should be diminished in an exponentially growing

population because the afflicted cells cannot expand their numbers. Conversely, if all cells undergo cytokinesis failure, centriole number would change predictably depending on the rules of centriole overproduction. It would definitely be pertinent to consider what could happen if cytokinesis failure conferred either a fitness advantage or changed the rules of centriole overproduction but the diversity of possible scenarios is so rich that it would merit its own, separate, investigation.

Implications for cancer evolution

Whether or not extra centrioles can lead to tumourigenesis remains a controversial topic in the field. The diversity of studies and disparaging results seem to indicate that the causal link between centriole number abnormalities and cancer formation is context-dependent at best (Coelho et al., 2015; Kulukian et al., 2015; Levine et al., 2017; Serçin et al., 2016; Shoshani et al., 2021; Vitre et al., 2015).

It has been proposed that low numbers of extra centrioles can be tumourigenic whereas high numbers might be too deleterious to allow cancer to develop (Bloomfield and Cimini, 2023; Levine et al., 2017). However, cells with low numbers of extra centrioles still suffer a significant fitness cost (Levine et al., 2017). Furthermore, our previous work supports the idea that selection is insensitive to the absolute number of extra centrioles (Dias Louro et al., 2021). Instead, the most important statistic to consider might be the relative frequency of cells carrying extra centrioles rather than their absolute, or mean, numbers.

Whereas we observed that empirical distributions of centriole numbers per cell at mitosis tend to be heavy-tailed, this was not true for all analysed data sets (Dias Louro et al., 2021). With this model, we observed light-tailed distributions when only a single pair of centrioles was allowed to duplicate. When all preexisting centrioles were allowed to duplicate, we obtained heavy-tailed

distributions. Both scenarios are probably unrealistic but they suggest that the fidelity of extra centriole duplication can control the tailedness of the distribution. If *most* extra centrioles are capable of duplication, the distributions should me more heavy-tailed, and vice-versa if fewer extra centrioles are capable of duplication. Experimental observations seem to indicate that the former is likelier (Marshall, 2007; Sala et al., 2020), but these were conducted in a limited number of systems. Based on our model predictions, it would be interesting to test if the cell lines in which heavy-tailed distributions were not observed duplicate extra centrioles less efficiently than others.

There is some evidence suggesting that asymmetric segregation of extra centrioles can improve the survival of daughter cells inheriting wild-type centriole numbers (Baudoin et al., 2020). Thus, asymmetric divisions could provide an elegant solution for coping with centrosome amplification. Our model supports this hypothesis if selection acts predominantly in G1 and the rate of centriole overproduction is high. If these conditions are met, cells might be under pressure to segregate centrioles asymmetrically. If not, we predict there is no particular advantage of predominantly asymmetric divisions over random centriole segregation. Existing data on centriole segregation was obtained in cell populations following experimental induction of centrosome amplification (Sala et al., 2020; Wang et al., 2011). It would be interesting to test if asymmetric divisions are more frequent in systems where cells with high centriole numbers are naturally observed and to precisely time when selection acts.

Furthermore, our models predict that asymmetric divisions are likelier to occur in cells with low numbers of extra centrioles. Asymmetric segregation of centrioles can favour chromosome missegregation (Cosenza et al., 2017). Thus, it is possible that cells with low numbers of extra centrioles optimise the probability of generating viable aneuploid progeny. Taken together, our models

seem to suggest that tumour development may be favoured for low centriole overproduction rates - yielding few cells with centrosome amplification, which impart a modest fitness cost for the population, and low numbers of extra centrioles, which promote asymmetric divisions. Indeed, live tumours and those generated following Plk4 overexpression tend to possess a low number of cells containing only a few extra centrioles (Levine et al., 2017). Thus, our results are coherent with current models of tumourigenesis initiated by centrosome amplification. However, it should be stressed that these models are still heavily debated.

Exploring the evolutionary dynamics of extra centrioles in the context of cancer formation was beyond the subject of this work. Ultimately, a definitive answer for the link between extra centrioles and cancer remains elusive. Regardless, our work shed new light on the conditions in which this *might* happen and hopefully can provide guidance for future research.

Interpreting centriole number abnormalities in interphase

It has been long established that screening centriole number abnormalities in interphase is subject to underscoring if the underlying cell cycle phase is not known (Bettencourt-Dias et al., 2011; Godinho and Pellman, 2014). While understanding the sources of underscoring might appear superfluous at first, if this underscoring was consistent, then it could be easily corrected when analysing data. Our modelling approach suggested that if one lacks the knowledge of *when* selection acts or the underlying distribution of cell cycle stages, this underscoring cannot be reliably corrected.

Scoring interphase cells does have practical advantages because mitotic cells tend to be relatively rare. In this regard, we could observe that, for low overproduction rates, which roughly correspond to low penetrance of centrosome amplification, empirically, scoring interphase cells did not lead to significant errors.

Thus, when screening low-abundance centriole number anomalies, opting for interphase cells might be adequate. Secondly, one might not need to know the precise stage of the cell cycle of each analysed cell. Rather, knowing the *distribution* of cell cycle stages, which can be readily attained by DNA staining and flow cytometry might be sufficient. Both these observations need further empirical validation.

The same can theoretically apply to tumour biopsies, although in that respect there are additional problems. Firstly, biopsies may section out significant portions of cells, such that centrioles might be left out altogether (Lopes et al., 2018; Wang et al., 2019). Secondly, the biology of live tissues is considerably more complex. In particular, exponential growth is unlikely to reflect the underlying proliferation dynamics - many cells may be quiescent, i.e. non-dividing, a biopsy may include multiple cell types, the 3D environment might affect cell division. Quantifying and properly assessing centriole number anomalies in patient samples thus requires correcting for these additional factors. Encouragingly, there are examples in which these abnormalities were found to be proportional to the ones found in representative cell lines (Lopes et al., 2018). Overall, a more detailed understanding of tumour cell biology might be necessary to correctly evaluate centriole number abnormalities, which may be crucial for prognosis. It would be interesting to assess how these factors could be accounted for to improve cancer prognosis.

Methods

The model described in this paper was implemented in Python as an individual-based stochastic model, using the package *gillespy2*, and a system of ordinary differential equations, in conjunction with the associated transition rate matrix. Numeric solutions were obtained by simulating the individual-based model and

integrating the system of ordinary differential equations, and compared to the computed dominant eigenvalues/eigenvectors. After verifying that the three implementations produced the same results, all analyses of the paper were conducted by finding the dominant eigenvalues and eigenvectors of the transition rate matrix using built-in functions.

Supplementary material

Preliminary analyses

Before conducting our analyses, we addressed the limitation of considering a finite maximum number of extra centrioles per cell, i_{max}. High centriole overproduction rates, β or low strength unspecific selection, w, could, for instance generate cells with extremely high centriole numbers. If i_{max} was not sufficiently high, it could generate artefacts in the distributions and lead to erroneous interpretations of model behaviour. We generated equilibrium distributions for extreme values of these parameters, as well as the inheritance probability α, while varying i_{max}. We observed that indeed, for low values of i_{max} ($i_{max} = 20$), the distributions looked considerably different from higher values. However, there was little difference in the distributions for higher values ($i_{max} = 60, 100$). Thus, for all subsequent analyses, we set the maximum number of extra centrioles per cell, i_{max} to 100.

Detailed model formulation

To extend our previous formalism on centriole number dynamics in proliferating cell populations, we developed a new model that accounted for cell cycle progression and explicitly differentiated between centriole biogenesis and inheritance of centrioles post-mitosis.

Our subject of focus was a population of cycling cells growing in an unlimited environment. We assume each cell was characterised by their current cell cycle stage, and the number of extra centrioles they harboured, at time t. As with previous modelling efforts, we assumed that cells could progress from G1, through to S, to G2, and M (mitosis). We assumed that centriole duplication occurred at least once and only once instantly when cells transitioned from G1 to S and that each daughter cell inherited at least two centrioles after mitosis. Thus, it was sufficient to track the number of extra centrioles, $i = 0, 1, 2...$ it contained.

To parameterise cell cycle progression, we assumed that G1, S, and G2 and M followed exponentially-distributed waiting times with mean τ_{G1}, τ_S, τ_{G2}, and τ_M, respectively, resulting in hypo-exponentially/gamma-distributed generation times. Other distributions (e.g. delayed exponential) have been used to model the waiting times of each cell cycle phase (Weber et al., 2014); we opted for exponential distributions because they offered an easily tractable solution which was adequately suited for our study. In the model, a cell divided after exiting M, yielding two daughter cells in G1. For simplification, we did not consider the possibility of a cell undergoing a multipolar division or failing to divide.

Although centriole assembly is yet to be precisely timed, it is generally accepted that centriolar precursors accumulate near the walls of preexisting centrioles during the G1-S transition, with biogenesis continuing through S-phase. The process of establishing a centriole assembly site and seeding of a new procentriole typically is negligibly small when compared to the length of the cell cycle. Therefore, we made the simplifying assumption that centriole biogenesis occurred instantaneously and simultaneously when cells transitioned from G1 to S. It is widely regarded that canonical centriole duplication occurs with extremely high fidelity. There is some data indicating that extra centrioles can also duplicate with comparable degrees of efficiency (Marshall,

2007; Sala et al., 2020). Thus, we assumed that each centriole in a G1 was at least duplicated once and only once. Overexpression of key molecules that trigger centriole assembly, such as Plk4, STIL, and SAS-6, is sufficient to induce overproduction of centrioles (Arquint et al., 2012; Bettencourt-Dias et al., 2005; Kleylein-Sohn et al., 2007; Strnad et al., 2007). Although the kinetics of *de novo* centriole assembly have been recently explored (Nabais et al., 2021, the kinetics of canonical centriole biogenesis (i.e. in close proximity to preexisting centrioles) remain poorly understood. As a simplifying assumption, we posited that extra centrioles could be produced following a Poisson distribution with mean β, which is equal to the centriole overproduction rate. This means that each centriole biogenesis event is independent of the previous one, implying no component limitations within the cell/centrosome or any spatiotemporal restrictions. This assumption is justified for the following reasons: first, there does not seem to exist a limiting cytoplasmic pool for several centriole duplication factors *Drosophila* embryos (Steinacker et al., n.d.). In human cells, centrioles can reduplicate after the procentriole is laser-ablated, could indirectly indicate that centrosomal components are not limited (Uetake et al., 2007). Second, the cell is many orders of magnitude larger than a centriole, such that significant spatial constraints are unlikely to affect centriole biogenesis, unless in extreme cases such as in multicilliated cells (Mercey et al., 2019). Even in the case of what is commonly referred to as centriole overduplication, at least seven procentrioles can emerge in a rosette-like configuration around a preexisting one (Lopes et al., 2018); so, even local spatial constraints may only be relevant in extreme cases of centriole overproduction. Third, and as previously mentioned, the duration of the cell cycle is considerably longer than the stage the cell is permissive to centriole biogenesis, such that cell cycle length is unlikely to affect centriole overproduction. Thus, in the absence of extreme rates of centriole overproduction, Poisson-distributed centriole assembly

events provided us with a straightforward implementation of this process with no need for currently unverifiable assumptions on its biophysical underpinnings. The rate at which a cell containing i extra centrioles generated a cell with j extra centrioles, assuming perfect duplication of all extra centrioles, is given by:

$$B_{i,j} = \frac{e^{-\beta}\beta^{(j-2i)}}{(j-2i)!} \tag{3.1}$$

As with canonical centriole duplication, centriole segregation in cells with wild-type centriole numbers is nearly deterministic. Segregation of extra centrioles, on the other hand, has been described as being random (Nakamura et al., 2013; Sala et al., 2020). Indeed, it has been reported that the distribution of differences in centriole number between daughter cells as a function of the number of centrioles in the mother cell was approximately Poisson-distributed, i.e. each extra centriole is independently segregated from the others (Sala et al., 2020). Based on these observations, and since our population was sub-divided into categories with finite centriole numbers, we assumed that centriole segregation followed a binomial distribution, such that any of the i centrioles in the mother cell can be inherited by one of the daughter cells with probability α.

Thus, both daughter cells will inherit i extra centrioles according to:

$$A_{i,j} = \binom{j}{i}\alpha^i(1-\alpha)^{j-i} + \binom{j}{i}(1-\alpha)^i\alpha^{j-i} \tag{3.2}$$

As the last component of our model, we implemented selection against extra centrioles. Multiple studies have showed that extra centrioles can efficiently trigger cell cycle delay during interphase

or mitosis, cell cycle arrest, and/or apoptosis (Coelho et al., 2015; Ganem et al., 2009; Holland et al., 2012; Kulukian et al., 2015; Sala et al., 2020; Serçin et al., 2016; Vitre et al., 2015). Moreover, our previous work suggested that these mechanisms may be equally efficient regardless of the number of extra centrioles (Dias Louro et al., 2021). We assumed that cells with i extra centrioles experienced a cell cycle delay w_i^{G1}, w_i^S, w_i^{G2}, and w_i^M in the corresponding cell cycle stage. We assumed that $\{w_i^{G1}$, w_i^S, w_i^{G2}, $w_i^M\}$={0,0,0,0} if $i = 0$ (i.e. if a cell harbours no extra centrioles); if $i > 0$, $\{w_i^{G1}$, w_i^S, w_i^{G2}, $w_i^M\}$=$\{w^{G1}$, w^S, w_i^{G2},$w_i^M\}$.

The rate of change in the population size for each sub-population, N_i^{G1}, N_i^S,N_i^{G2},N_i^M, containing a certain number of extra centrioles, i, in either of the four cell cycle stages, can thus be approximated by the following system of ordinary differential equations.

$$\frac{\partial N_i^{G1}}{\partial t} = \sum_{j=i} A_{i,j} \frac{1}{\tau_M}(1 - w_i^M) N_j^M(t) - \frac{1}{\tau_{G1}}(1 - w_i^{G1}) N_i^{G1}(t) \tag{3.3}$$

$$\frac{\partial N_j^S}{\partial t} = \sum_{i=0}^{j} B_{i,j} \frac{1}{\tau_{G1}}(1 - w_i^{G1}) N_i^{G1}(t) - \frac{1}{\tau_S}(1 - w_i^S) N_j^S(t) \tag{3.4}$$

$$\frac{\partial N_j^{G2}}{\partial t} = \frac{1}{\tau_S}(1 - w_i^S) N_i^S(t) - \frac{1}{\tau_{G2}}(1 - w_j^{G2}) N_i^{G2}(t) \tag{3.5}$$

$$\frac{\partial N_j^M}{\partial t} = \frac{1}{\tau_{G2}}(1 - w^G 2_i) N_i^{G2}(t) - \frac{1}{\tau_M}(1 - w_j^M) N_i^M(t) \tag{3.6}$$

The relative frequencies of each sub-population at time t can then be obtained by dividing their population size by the total population size $\sum_i N_i^{G1}$+N_i^S+N_i^{G2}+N_i^M. As previously mentioned, for practical reasons, we assumed that a cell can contain a maximum number of centrioles, i_{max}.

Our model was similar to previous models of the eukaryotic cell cycle, with the addition of an arbitrary number of centriole

number classes. Such models cannot be solved analytically but can be solved numerically. Notwithstanding, it can be shown they tend asymptotically towards "balanced exponential growth", i.e. an exponentially growing population in which the relative abundances of cells in each cell cycle stage are at a stable equilibrium. We were particularly interested in these conditions as an extension of the aforementioned centriole overproduction-selection balance. With that in mind, the exponential growth rate and equilibrium frequencies of each sub-population could be obtained by calculating the highest real-valued eigenvalue and corresponding normalised eigenvector of the transition rate matrix associated with the system of ordinary differential equations above. We performed these calculations numerically as described in the Methods sections.

Supplementary figures

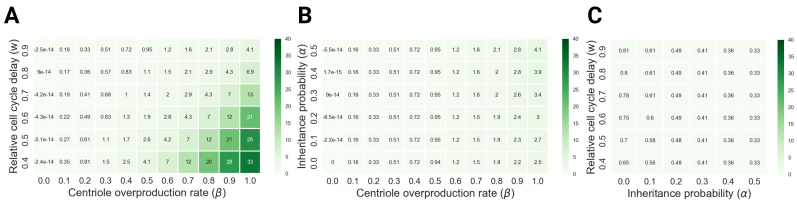


FIGURE S1: **Mean equilibrium centriole number in mitotic cells as a function of model parameters under unspecific selection.** Bi-parametric heatmaps of mean centriole numbers at equilibrium in mitotic cells as a function of each pair of model parameters. A - Centriole overproduction rate versus unspecific relative cell cycle delay; B - Centriole overproduction rate versus inheritance probability; C - Inheritance probability versus unspecific relative cell cycle delay. The default parameter values were $\beta = 0.2$, $w = 0.9$, $\alpha = 0.5$.

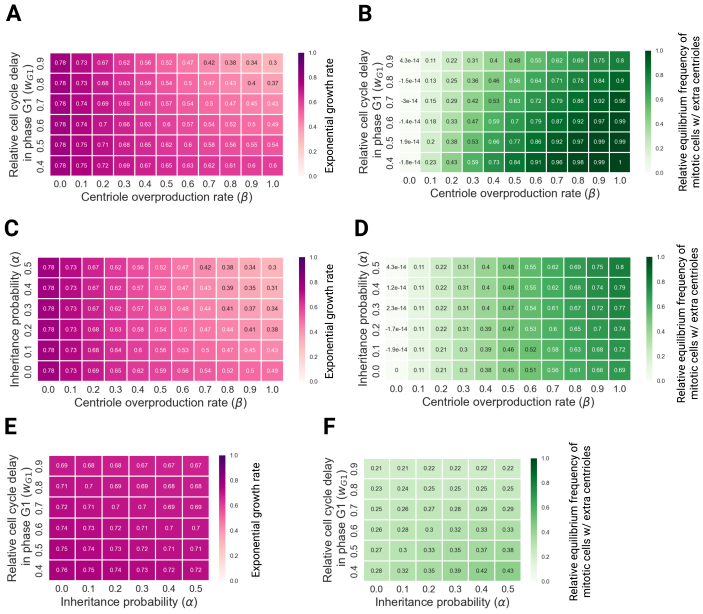


FIGURE S2: **Exponential growth rate and relative frequency of mitotic cells as a function of model parameters when selection acts specifically in G1.** Biparametric heatmaps of the exponential growth rate (A, C, E) and relative frequency of mitotic cells with extra centrioles at equilibrium (B, D, F) as a function of each pair of parameters. A, B - Centriole overproduction rate versus G1-specific relative cell cycle delay; C, D - Centriole overproduction rate versus inheritance probability; E, F - Inheritance probability versus G1-specific relative cell cycle delay. The default parameter values were $\beta = 0.2$, $w = 0.9$, $\alpha = 0.5$.

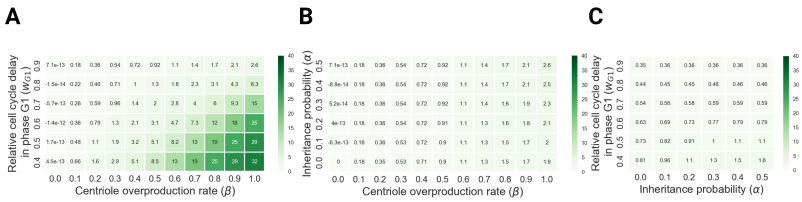


FIGURE S3: **Mean equilibrium centriole number in mitotic cells as a function of model parameters when selection acts specifically in G1.** Biparametric heatmaps of mean centriole numbers at equilibrium in mitotic cells as a function of each pair of model parameters. A - Centriole overproduction rate versus G1-specific relative cell cycle delay; B - Centriole overproduction rate versus inheritance probability; C - Inheritance probability versus G1-specific relative cell cycle delay. The default parameter values were $\beta = 0.2$, $w = 0.9$, $\alpha = 0.5$.

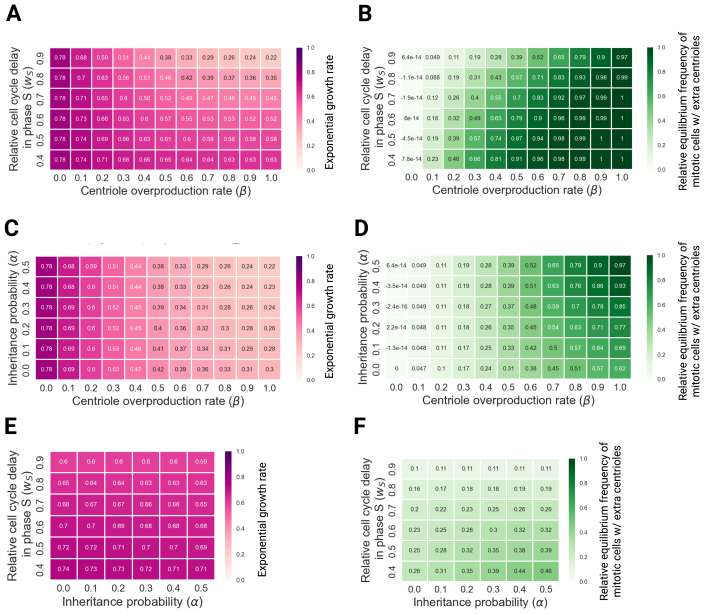


FIGURE S4: **Exponential growth rate and relative frequency of mitotic cells as a function of model parameters when selection acts specifically in S.** Biparametric heatmaps of the exponential growth rate (A, C, E) and relative frequency of mitotic cells with extra centrioles at equilibrium (B, D, F) as a function of each pair of parameters. A, B - Centriole overproduction rate versus S-specific relative cell cycle delay; C, D - Centriole overproduction rate versus inheritance probability; E, F - Inheritance probability versus S-specific relative cell cycle delay. The default parameter values were $\beta = 0.2$, $w = 0.9$, $\alpha = 0.5$.

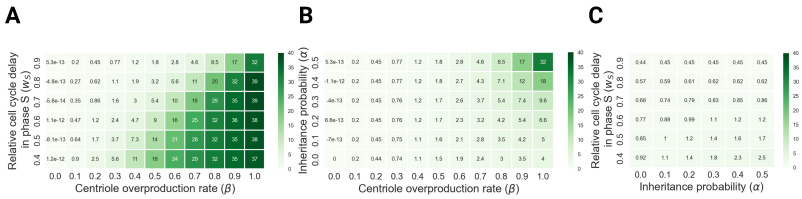


FIGURE S5: **Mean equilibrium centriole number in mitotic cells as a function of model parameters when selection acts specifically in S.** Biparametric heatmaps of mean centriole numbers at equilibrium in mitotic cells as a function of each pair of model parameters. A - Centriole overproduction rate versus S-specific relative cell cycle delay; B - Centriole overproduction rate versus inheritance probability; C - Inheritance probability versus S-specific relative cell cycle delay. The default parameter values were $\beta = 0.2$, $w = 0.9$, $\alpha = 0.5$.

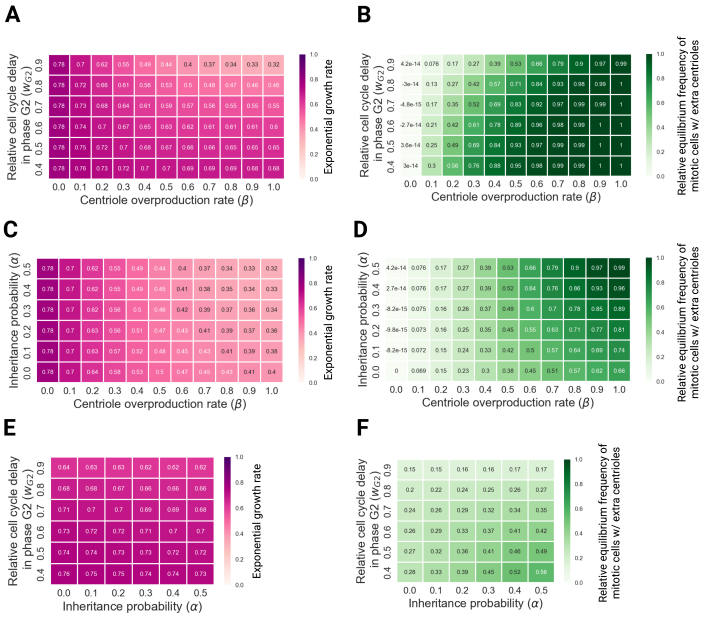


FIGURE S6: **Exponential growth rate and relative frequency of mitotic cells as a function of model parameters when selection acts specifically in G2.** Biparametric heatmaps of the exponential growth rate (A, C, E) and relative frequency of mitotic cells with extra centrioles at equilibrium (B, D, F) as a function of each pair of parameters. A, B - Centriole overproduction rate versus G2-specific relative cell cycle delay; C, D - Centriole overproduction rate versus inheritance probability; E, F - Inheritance probability versus G2-specific relative cell cycle delay. The default parameter values were $\beta = 0.2$, $w = 0.9$, $\alpha = 0.5$.

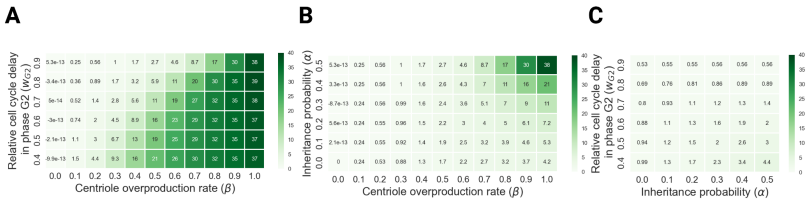


FIGURE S7: **Mean equilibrium centriole number in mitotic cells as a function of model parameters.** Biparametric heatmaps of mean centriole numbers at equilibrium in mitotic cells as a function of model parameters when selection acts specifically in G2. A - Centriole overproduction rate versus G2-specific relative cell cycle delay; B - Centriole overproduction rate versus inheritance probability; C - Inheritance probability versus G2-specific relative cell cycle delay. The default parameter values were $\beta = 0.2$, $w = 0.9$, $\alpha = 0.5$.

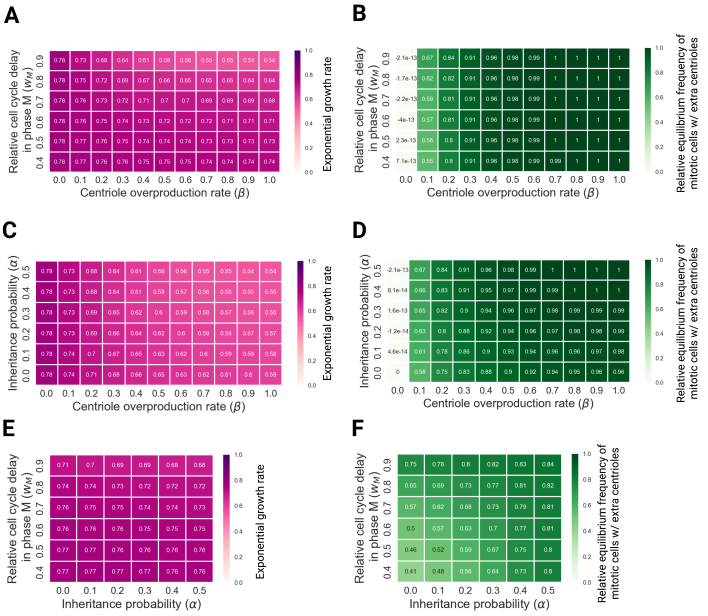


FIGURE S8: **Exponential growth rate and relative frequency of mitotic cells as a function of model parameters when selection acts specifically in M.** Biparametric heatmaps of the exponential growth rate (A, C, E) and relative frequency of mitotic cells with extra centrioles at equilibrium (B, D, F) as a function of each pair of parameters. A, B - Centriole overproduction rate versus M-specific relative cell cycle delay; C, D - Centriole overproduction rate versus inheritance probability; E, F - Inheritance probability versus M-specific relative cell cycle delay. The default parameter values were $\beta = 0.2$, $w = 0.9$, $\alpha = 0.5$.

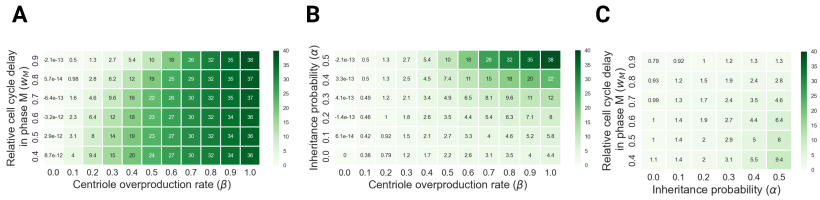


FIGURE S9: **Mean equilibrium centriole number in mitotic cells as a function of model parameters when selection acts specifically in M.** Biparametric heatmaps of mean centriole numbers at equilibrium in mitotic cells as a function of each pair of model parameters. A - Centriole overproduction rate versus M-specific relative cell cycle delay; B - Centriole overproduction rate versus inheritance probability; C - Inheritance probability versus M-specific relative cell cycle delay. The default parameter values were $\beta = 0.2$, $w = 0.9$, $\alpha = 0.5$.

References

Arquint, C., Sonnen, K. F., Stierhof, Y.-D., & Nigg, E. A. (2012). Cell-cycle-regulated expression of STIL controls centriole number in human cells. *Journal of Cell Science*, *125*(5), 1342–1352. https://doi.org/10.1242/jcs.099887

Baudoin, N. C., Nicholson, J. M., Soto, K., Martin, O., Chen, J., & Cimini, D. (2020). Asymmetric clustering of centrosomes defines the early evolution of tetraploid cells. *eLife*, *9*. https://doi.org/10.7554/eLife.54565

Bettencourt-Dias, M., Rodrigues-Martins, A., Carpenter, L., Riparbelli, M., Lehmann, L., Gatt, M. K., Carmo, N., Balloux, F., Callaini, G., & Glover, D. M. (2005). SAK/PLK4 is required for centriole duplication and flagella development. *Current Biology*, *15*(24), 2199–2207. https://doi.org/10.1016/j.cub.2005.11.042

Bettencourt-Dias, M., Hildebrandt, F., Pellman, D., Woods, G., & Godinho, S. A. (2011). Centrosomes and cilia in human disease. *Trends in Genetics*, *27*(8), 307–315. https://doi.org/10.1016/j.tig.2011.05.004

Bloomfield, M., & Cimini, D. (2023). The fate of extra centrosomes in newly formed tetraploid cells: Should i stay, or should i go? *Frontiers in Cell and Developmental Biology*, *11*. https://doi.org/10.3389/fcell.2023.1210983

Breslow, D. K., & Holland, A. J. (2019). Mechanism and Regulation of Centriole and Cilium Biogenesis. *Annual Review of Biochemistry*, *88*(1), 691–724. https://doi.org/10.1146/annurev-biochem-013118-111153

Brito, D. A., Gouveia, S. M., & Bettencourt-Dias, M. (2012). Deconstructing the centriole: Structure and number control. *Current Opinion in Cell Biology*, *24*(1), 4–13. https://doi.org/10.1016/j.ceb.2012.01.003

Burigotto, M., Mattivi, A., Migliorati, D., Magnani, G., Valentini, C., Roccuzzo, M., Offterdinger, M., Pizzato, M., Schmidt,

A., Villunger, A., Maffini, S., & Fava, L. L. (2021). Centriolar distal appendages activate the centrosome-piddosome-p53 signalling axis via ankrd26. *The EMBO Journal*, *40*(4), e104844. https://doi.org/https://doi.org/10.15252/embj.2020104844

Chan, J. Y. (2011). A clinical overview of centrosome amplification in human cancers. *Int. J. Biol. Sci.*, *7*(8), 1122–1144.

Coelho, P. A., Bury, L., Shahbazi, M. N., Liakath-Ali, K., Tate, P. H., Wormald, S., Hindley, C. J., Huch, M., Archer, J., Skarnes, W. C., Zernicka-Goetz, M., & Glover, D. M. (2015). Over-expression of Plk4 induces centrosome amplification, loss of primary cilia and associated tissue hyperplasia in the mouse. *Open Biology*, *5*(12). https://doi.org/10.1098/rsob.150209

Cosenza, M. R., Cazzola, A., Rossberg, A., Schieber, N. L., Konotop, G., Bausch, E., Slynko, A., Holland-Letz, T., Raab, M. S., Dubash, T., Glimm, H., Poppelreuther, S., Herold-Mende, C., Schwab, Y., & Krämer, A. (2017). Asymmetric centriole numbers at spindle poles cause chromosome missegregation in cancer. *Cell Reports*, *20*(8), 1906–1920. https://doi.org/https://doi.org/10.1016/j.celrep.2017.08.005

Dias Louro, M. A., Bettencourt-Dias, M., & Bank, C. (2021). Patterns of selection against centrosome amplification in human cell lines. *PLOS Computational Biology*, *17*(5), 1–22. https://doi.org/10.1371/journal.pcbi.1008765

Fava, L. L., Schuler, F., Sladky, V., Haschka, M. D., Soratroi, C., Eiterer, L., Demetz, E., Weiss, G., Geley, S., Nigg, E. A., & Villunger, A. (2017). The PIDDosome activates p53 in response to supernumerary centrosomes. *Genes and Development*, *31*(1), 34–45. https://doi.org/10.1101/gad.289728.116

Galofré, C., Asensio, E., Ubach, M., Torres, I. M., Quintanilla, I., Castells, A., & Camps, J. (2020). Centrosome reduction

in newly-generated tetraploid cancer cells obtained by separase depletion. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-65975-1

Ganem, N. J., Godinho, S. A., & Pellman, D. (2009). A mechanism linking extra centrosomes to chromosomal instability. *Nature*, *460*(7252), 278–282. https://doi.org/10.1038/nature08136

Godinho, S. A., & Pellman, D. (2014). Causes and consequences of centrosome abnormalities in cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1650). https://doi.org/10.1098/rstb.2013.0467

Gomes Pereira, S., Dias Louro, M. A., & Bettencourt-Dias, M. (2021). Biophysical and Quantitative Principles of Centrosome Biogenesis and Structure. *Nature Communications*, *37*, 43–63. https://doi.org/10.1038/s41467-017-00238-8.

Gönczy, P. (2015). Centrosomes and cancer: Revisiting a long-standing relationship. *Nature Reviews Cancer*, *15*(11), 639–652. https://doi.org/10.1038/nrc3995

Holland, A. J., Fachinetti, D., Zhu, Q., Bauer, M., Verma, I. M., Nigg, E. A., & Cleveland, D. W. (2012). The autoregulated instability of polo-like kinase 4 limits centrosome duplication to once per cell cycle. *Genes Dev.*, *26*(24), 2684–2689.

Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Habedanck, R., Stierhof, Y.-D., & Nigg, E. A. (2007). Plk4-induced centriole biogenesis in human cells. *Developmental Cell*, *13*(2), 190–202. https://doi.org/https://doi.org/10.1016/j.devcel.2007.07.002

Kulukian, A., Holland, A. J., Vitrec, B., Naika, S., Cleveland, D. W., & Fuchsa, E. (2015). Epidermal development, growth control, and homeostasis in the face of centrosome amplification. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(46), E6311–E6320. https://doi.org/10.1073/pnas.1518376112

Kwon, M., Godinho, S. A., Chandhok, N. S., Ganem, N. J., Azioune, A., Thery, M., & Pellman, D. (2008). Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes. *Genes and Development*, *22*(16), 2189–2203. https://doi.org/10.1101/gad.1700908

Lambrus, B. G., Uetake, Y., Clutario, K. M., Daggubati, V., Snyder, M., Sluder, G., & Holland, A. J. (2015). P53 protects against genome instability following centriole duplication failure. *Journal of Cell Biology*, *210*(1), 63–77. https://doi.org/10.1083/jcb.201502089

Lau, T. Y., & Poon, R. Y. (2023). Whole-genome duplication and genome instability in cancer cells: Double the trouble. *International Journal of Molecular Sciences*, *24*(4). https://doi.org/10.3390/ijms24043733

Levine, M. S., Bakker, B., Boeckx, B., Moyett, J., Lu, J., Vitre, B., Spierings, D. C., Lansdorp, P. M., Cleveland, D. W., Lambrechts, D., Foijer, F., & Holland, A. J. (2017). Centrosome Amplification Is Sufficient to Promote Spontaneous Tumorigenesis in Mammals. *Developmental Cell*, *40*(3), 313–322.e5. https://doi.org/10.1016/j.devcel.2016.12.022

Lopes, C. A. M., Jana, S. C., Cunha-Ferreira, I., Zitouni, S., Bento, I., Duarte, P., Gilberto, S., Freixo, F., Guerrero, A., Francia, M., Lince-Faria, M., Carneiro, J., & Bettencourt-Dias, M. (2015). PLK4 trans-Autoactivation Controls Centriole Biogenesis in Space. *Developmental Cell*, *35*(2), 222–235. https://doi.org/10.1016/j.devcel.2015.09.020

Lopes, C. A., Mesquita, M., Cunha, A. I., Cardoso, J., Carapeta, S., Laranjeira, C., Pinto, A. E., Pereira-Leal, J. B., Dias-Pereira, A., Bettencourt-Dias, M., & Chaves, P. (2018). Centrosome amplification arises before neoplasia and increases upon p53 loss in tumorigenesis. *Journal of Cell Biology*, *217*(7), 2353–2363. https://doi.org/10.1083/jcb.201711191

Marshall, W. F. (2007). Stability and robustness of an organelle number control system: modeling and measuring homeostatic regulation of centriole abundance. *Biophysical journal*, *93*(5), 1818–1833. https://doi.org/10.1529/biophysj.107.107052

Marshall, W. F. (2016). Cell Geometry: How Cells Count and Measure Size. *Annual Review of Biophysics*, *45*(1), 49–64. https://doi.org/10.1146/annurev-biophys-062215-010905

Marteil, G., Guerrero, A., Vieira, A. F., De Almeida, B. P., Machado, P., Mendonça, S., Mesquita, M., Villarreal, B., Fonseca, I., Francia, M. E., Dores, K., Martins, N. P., Jana, S. C., Tranfield, E. M., Barbosa-Morais, N. L., Paredes, J., Pellman, D., Godinho, S. A., & Bettencourt-Dias, M. (2018). Overelongation of centrioles in cancer promotes centriole amplification and chromosome missegregation. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-03641-x

Mercey, O., Levine, M. S., LoMastro, G. M., Rostaing, P., Brotslaw, E., Gomez, V., Kumar, A., Spassky, N., Mitchell, B. J., Meunier, A., & Holland, A. J. (2019). Massive centriole production can occur in the absence of deuterosomes in multiciliated cells. *Nature Cell Biology*, *21*, 1544–1552. https://doi.org/10.1038/s41556-019-0427-x

Mukherji, S., & O'Shea, E. K. (2014). Mechanisms of organelle biogenesis govern stochastic fluctuations in organelle abundance. *eLife*, *2014*(3). https://doi.org/10.7554/eLife.02678.001

Nabais, C., Pessoa, D., de Carvalho, J., van Zanten, T., Duarte, P., Mayor, S., Carneiro, J., Telley, I. A., & Bettencourt-Dias, M. (2021). Plk4 triggers autonomous de novo centriole biogenesis and maturation. *Journal of Cell Biology*, *220*(5), e202008090. https://doi.org/10.1083/jcb.202008090

Nakamura, T., Saito, H., & Takekawa, M. (2013). Sapk pathways and p53 cooperatively regulate plk4 activity and centrosome integrity under stress. *Nature Communications*, *4*(1), 1775. https://doi.org/10.1038/ncomms2752

Ritter, M. A., Fowler, J. F., Kim, Y. J., Gilchrist, K. W., Morrissey, L. W., & Kinsella, T. J. (1994). Tumor cell kinetics using two labels and flow cytometry. *Cytometry*, *16*(1), 49–58. https://doi.org/https://doi.org/10.1002/cyto.990160108

Sala, R., Farrell, K., & Stearns, T. (2020). Growth disadvantage associated with centrosome amplification drives population-level centriole number homeostasis. *Molecular Biology of the Cell*, mbc.E19–04–0195. https://doi.org/10.1091/mbc.e19-04-0195

Serçin, Ö., Larsimont, J. C., Karambelas, A. E., Marthiens, V., Moers, V., Boeckx, B., Le Mercier, M., Lambrechts, D., Basto, R., & Blanpain, C. (2016). Transient PLK4 overexpression accelerates tumorigenesis in p53-deficient epidermis. *Nature Cell Biology*, *18*(1), 100–110. https://doi.org/10.1038/ncb3270

Shoshani, O., Bakker, B., de Haan, L., Tijhuis, A. E., Wang, Y., Kim, D. H., Maldonado, M., Demarest, M. A., Artates, J., Zhengyu, O., Mark, A., Wardenaar, R., Sasik, R., Spierings, D. C. J., Vitre, B., Fisch, K., Foijer, F., & Cleveland, D. W. (2021). Transient genomic instability drives tumorigenesis through accelerated clonal evolution. *Genes Dev.*, *35*(15-16), 1093–1108.

Steinacker, T., Wong, S, Novak, Z., Saurya, S, Gartenmann, L, van Houtum, E., Sayers, J., Lagerholm, B., & Raff, J. (n.d.). Centriole growth is not limited by a finite pool of components, but is limited by the cdk1/cyclin-dependent phosphorylation of ana2/sti. https://europepmc.org/article/PPR/PPR455086

Strnad, P., Leidel, S., Vinogradova, T., Euteneuer, U., Khodjakov, A., & Gönczy, P. (2007). Regulated HsSAS-6 Levels Ensure Formation of a Single Procentriole per Centriole during the Centrosome Duplication Cycle. *Developmental Cell*, *13*(2), 203–213. https://doi.org/10.1016/j.devcel.2007.07.004

Tkach, J. M., Philip, R., Sharma, A., Strecker, J., Durocher, D., & Pelletier, L. (2022). Global cellular response to chemical perturbation of plk4 activity and abnormal centrosome number (J. Lüders & A. Akhmanova, Eds.). *eLife*, *11*, e73944. https://doi.org/10.7554/eLife.73944

Uetake, Y., Loncarek, J., Nordberg, J. J., English, C. N., La Terra, S., Khodjakov, A., & Sluder, G. (2007). Cell cycle progression and de novo centriole assembly after centrosomal removal in untransformed human cells. *Journal of Cell Biology*, *176*(2), 173–182. https://doi.org/10.1083/jcb.200607073

Vitre, B., Holland, A. J., Kulukian, A., Shoshani, O., Hirai, M., Wanga, Y., Maldonado, M., Cho, T., Boubaker, J., Swing, D. A., Tessarollo, L., Evans, S. M., Fuchs, E., & Cleveland, D. W. (2015). Chronic centrosome amplification without tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(46), E6321–E6330. https://doi.org/10.1073/pnas.1519388112

Wang, M., Knudsen, B. S., Nagle, R. B., Rogers, G. C., & Cress, A. E. (2019). A method of quantifying centrosomes at the single-cell level in human normal and cancer tissue. *Molecular Biology of the Cell*, *30*(7), 811–819. https://doi.org/10.1091/mbc.E18-10-0651

Wang, W.-J., Soni, R. K., Uryu, K., & Bryan Tsou, M.-F. (2011). The conversion of centrioles to centrosomes: essential coupling of duplication with segregation. *Journal of Cell Biology*, *193*(4), 727–739. https://doi.org/10.1083/jcb.201101109

Weber, T. S., Jaehnert, I., Schichor, C., Or-Guil, M., & Carneiro, J. (2014). Quantifying the length and variance of the eukaryotic cell cycle phases by a stochastic model and dual nucleoside pulse labelling. *PLOS Computational Biology*, *10*(7), 1–17. https://doi.org/10.1371/journal.pcbi.1003616

Wong, Y. L., Anzola, J. V., Davis, R. L., Yoon, M., Motamedi, A., Kroll, A., Seo, C. P., Hsia, J. E., Kim, S. K., Mitchell, J. W., Mitchell, B. J., Desai, A., Gahman, T. C., Shiau, A. K., & Oegema, K. (2015). Reversible centriole depletion with an inhibitor of Polo-like kinase 4. *Science*, *348*(6239), 1155–1160. https://doi.org/10.1126/science.aaa5111

Yanagisawa, M., Dolbeare, F., Todoroki, T., & Gray, J. W. (1985). Cell cycle analysis using numerical simulation of bivariate dna/bromodeoxyuridine distributions. *Cytometry*, *6*(6), 550–562. https://doi.org/https://doi.org/10.1002/cyto.990060609

Chapter 4

Short-term evolution of cells with centrosome amplification

Marco António Dias Louro[1,2,3*], Catarina Peneda[1,4*], Claudia Bank[2,3], Mónica Bettencourt-Dias[1]

1. Instituto Gulbenkian de Ciência, Oeiras, Portugal

2. Institut für Ökologie und Evolution, Universität Bern, Bern, Switzerland

3. Swiss Institute of Bioinformatics, Lausanne, Switzerland

4. Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal

∗ These authors contributed equally to this work

Author contributions

This chapter is a preliminary draft for which we are still conducting some experiments and analyses. We expect to submit it soon. The contributions of each author for this work were as follows: Marco António Dias Louro (the candidate), Catarina Peneda, Claudia Bank, and Mónica Bettencourt-Dias conceptualised this study. Marco António Dias Louro performed cell culture with the help of Catarina Peneda, and all immunostainings. Marco António Dias Louro and Catarina Peneda designed all experiments, performed the qPCRs and competition assays, and performed epifluorescence microscopy image acquisition. Marco António Dias Louro performed all flow cytometry and image analyses, produced the plots and figures, and wrote this draft. Catarina Peneda, Claudia Bank, and Mónica Bettencourt-Dias edited the manuscript. Funding for this project was awarded to Marco António Dias Louro and Mónica Bettencourt-Dias. Marco António Dias Louro and Catarina Peneda contributed equally for to this work.

Abstract

The population-level processes that regulate the abundance of subcellular structures and their physiological consequences remain poorly understood. Centrioles are diminutive structures that organise the centrosome in various eukaryotes. The number of centrioles in healthy proliferating cells is usually kept constant through a tightly regulated duplication and segregation cycle. Centrosome amplification, or the gain of an excess number of centrioles, tends to be negatively selected but is commonly observed in cancer. It has been proposed that a modest increase in centriole numbers can be pro-tumourigenic whereas severe centrosome amplification is too deleterious to promote cancer development. However, how different degrees of centrosome amplification affect the physiology and evolution of proliferating cells remains undisclosed.

In this study we asked how populations of non-transformed cells respond when centriole numbers were chronically perturbed to different extents. Making use of non-transformed breast epithelial cell line where the levels of Plk4, the master regulator of centriole biogenesis, could be manipulated in a dose-dependent way through the addition of doxycycline, we devised a long-term cell culturing setup to answer this question and tracked a variety of cell physiological parameters over two months. Initially, we observed a dose-dependent increase in the relative frequency of cells with extra centrioles following the addition of doxycycline, which steadily decreased over time. Eventually, these cell populations approached basal levels of centrosome amplification but the rate of decrease in the relative frequency of cells with extra centrioles was positively correlated with doxycycline concentration. In addition, we found that relative fitness of these populations increased over time and was negatively correlated with the level of centrosome amplification. The fitness cost associated with abnormally high centriole numbers was caused by

slower proliferation and/or cell cycle arrest but not by increased cell death. Plk4 mRNA levels in doxycycline-treated populations were initially elevated before experiencing a transient decrease and returning to the original degree of overexpression. Similarly, the fluorescence signal intensity of centrosomal Plk4 increased following doxycycline addition but was indistinguishable from non-induced populations after one and two months. These results suggest a complex regulatory response that allowed cells to adapt to chronic Plk4 overexpression by suppressing centriole overproduction transcriptionally and post-transcriptionally. Taken together, our results suggest that chronic centrosome amplification is countered by negative selection and by inhibition of Plk4-mediated centriole overproduction.

Keywords: centrosome amplification, cell culture, fitness, Plk4

Introduction

Quantifying the forces that shape the abundances of subcellular components is a fundamental problem in cell biology. Whereas a significant body of research has delved into the biochemical underpinnings of biosynthetic control of organelle numbers (Karsenti, 2008), how population-level processes affect the abundance of cellular structures remains poorly understood. This aspect is particularly relevant because variation in the numbers of subcelullar structure is a source of phenotypic heterogeneity between cells, which can be targeted by selection. In addition, organelle copy number changes have been documented in a variety of tumour types and play key roles in altering cancer cell physiology (Vyas et al., 2016) Thus, dissecting the causes and consequences of variation in the numbers of subcellular structures is determinant for understanding cancer development.

In many proliferating cells of vertebrates and other eukaryotes,

centrioles are the core structures of the main microtubule organising centre, the centrosome. During the G1 phase of the cell cycle, most healthy cells contain two centrioles encased in a complex matrix of proteins called the pericentriolar material. During S-phase, the two original centrioles duplicate once and only once through a biosynthetic process that is rate-limited by a small number of key proteins, such as Plk4, STIL, and SAS-6 (Arquint et al., 2012; Cunha-Ferreira et al., 2013; Cunha-Ferreira et al., 2009; Lopes et al., 2015; Ohta et al., 2014; Ohta et al., 2018; Strnad et al., 2007; Takao et al., 2019a; Takao et al., 2019b). In late G2, the two centriole pairs, each containing a recently duplicated procentriole and its associated mother, separate and begin migrating to opposite poles of the cell, generating two independent centrosomes. During mitosis, each centrosome directs the organisation of the mitotic spindle and facilitates equal segregation of genetic material. Following cytokinesis, both daughter cells inherit a single centrosome with two centrioles, thus concluding the cycle (Breslow and Holland, 2019; Brito et al., 2012; Gomes Pereira et al., 2021; Gönczy, 2015).

The canonical centriole duplication and segregation cycle can be maintained by two synergistic regulatory mechanisms. First, centriole biogenesis is tightly regulated and cells rarely generate an "incorrect" number of centrioles. Secondly, cells with aberrant centriole numbers are heavily counter-selected. The expansion of cells with extra centrioles can be limited via the induction of cell death following aneuploid multipolar or pseudo-bipolar divisions (Kwon et al., 2008) or cell cycle arrest/delay (Fava et al., 2017; Holland et al., 2012; Sala et al., 2020). These processes reduce the relative growth rate of such cells in the population and limit the propagation of numerical errors in centrioles. Thus, regulation of biogenesis and selective elimination of cells carrying centriole number anomalies allows centriole number to be kept constant across cell divisions.

More generally, the percentage of cells carrying centriole number anomalies in populations of human cell lines tends to be idiosyncratic (Bettencourt-Dias et al., 2011), suggesting that centriole numbers kept at equilibrium. There are several lines of evidence suggesting that the original distribution of centriole numbers per cell is recovered upon transient perturbation, for example, through induction of cytokinesis failure (Baudoin et al., 2020; Galofré et al., 2020, overexpression of Plk4 (Sala et al., 2020), or inactivation/degradation of Plk4 (Lambrus et al., 2015; Wong et al., 2015), providing that the cells can continue proliferating. Therefore, centriole numbers appear to be maintained at an equilibrium between centriole overproduction and negative selection against cells with extra centrioles.

A quantitative understanding of this purported overproduction-selection balance is paramount for unravelling the still contentious role of extra centrioles, or centrosome amplification, in cancer development. Whether or not extra centrioles facilitate tumourigenesis seems to be highly context-dependent (Coelho et al., 2015; Kulukian et al., 2015; Levine et al., 2017; Serçin et al., 2016; Shoshani et al., 2021; Vitre et al., 2015). It has been proposed that modest levels of (transient or chronic) Plk4 overexpression can promote the appearance of certain cancer types whereas high levels lead to generalised tissue anomalies without malignant transformation (Levine et al., 2017). If this hypothesis is correct, then one could expect that: 1) relative fitness of a population of cells decreases as a function of the frequency of cells with extra centrioles/mean centriole number and/or; 2) populations exposed to a modest centriole number increase become more cancer-like over time compared to populations with canonical or extremely high numbers of centrioles. However, the how the level of Plk4 overexpression and, consequently, the degree of centrosome amplification in the population, affect cell physiology or evolution was yet to be tested.

In this study we asked how populations of non-transformed human cells respond to different levels of chronic Plk4 overexpression. Making use of a breast epithelial cell line carrying a doxycycline-inducible Plk4 transgene (Godinho et al., 2014), we cultured cell populations in the presence of different concentrations of doxycycline for over two months. Cells initially experienced centrosome amplification at different penetrances, proportional to the dose of doxycycline. However cells with extra centrioles gradually decreased in frequency and returned to their original numbers. We observed that cell populations bearing extra centrioles suffered a fitness penalty depending on level of centrosome amplification. This fitness penalty was caused by reduced proliferation and cell cycle delay/arrest. As wild-type centriole numbers were recovered, cell populations recovered their initial growth rate. In addition, we observed continuous Plk4 overexpression, despite a transient reduction in mRNA levels, and a decrease in centrosomal Plk4 levels. These results suggest that centriole overproduction was suppressed by a complex transcriptional and/or post-transcriptional regulation of Plk4 levels. We concluded that negative selection and down-regulation of Plk4 prevent extra centrioles from being maintained in proliferating cell populations, the dynamics of which were positively correlated with the initial levels of centrosome amplification.

Results

Setup to investigate the evolution of cells with extra centrioles

We were interested in addressing how cell populations respond when centriole numbers are chronically perturbed. To answer this question we devised a cell-culturing setup using the non-transformed human breast cell line MCF10A-Plk4 as a model

system (Godinho et al., 2014). Our choice was based on certain key properties of this cell line: 1) MCF10A-Plk4 allows experimental manipulation of centriole numbers via overexpression of a doxycycline-inducible Plk4 transgene; 2) the effects of doxycycline-induced Plk4 overexpression *per se* can be partially controlled by using a related cell line, MCF10A-Plk4$^{1-608}$, carrying a truncated form of Plk4. Overexpression of the truncated Plk4 does not affect centriole number as it lacks a critical part of the protein required for its centrosomal localisation (Guderian et al., 2010); 3) the effects of doxycycline treatment *per se* can be controlled for by using the parental cell line of MCF10A-Plk4 and MCF10A-Plk4$^{1-608}$, MCF10A-TetR, which does not contain the target Plk4 transgene (Godinho et al., 2014).

Treating MCF10A-Plk4 with doxycycline at a concentration of 2 μg/mL for 24-72h results in approximately 70% of cells with extra centrioles at mitosis, compared to approximately 10% in non-induced populations (Godinho et al., 2014). To test if we could obtain populations with different initial levels of centrosome amplification, we grew MCF10A-Plk4 in cell culture medium supplemented with doxycycline at a range of concentrations from 0 to 0.1 μg/mL for 24h. We manually scored the number of centrioles in mitotic cells by immunostaining based on the co-localisation of two centriolar proteins, centrin and Cep135 (1A). Mitotic cells were visually identified based on DNA staining. Analysing mitotic cells allowed us to more accurately estimate the degree of centriole number abnormalities, since healthy mitotic cells are expected to have four centrioles. Interphase cells, on the other hand, typically harbour two to four centrioles. Thus, only interphase cells with more than four centrioles can be accurately assessed as abnormal, which leads to their underscoring.

We observed that the relative frequency of mitotic cells with extra centrioles increased in an approximately monotonic fashion with the dose of doxycycline, eventually reaching a plateau 1B.

Our results indicated significant differences between populations grown with doxycycline at 0 μg/mL compared with and 0.001 or 0.1 μg/mL (relative frequency of mitotic cells with extra centrioles: 0 μg/mL - 0.0786 $\pm$ 0.040; 0.001 μg/mL - 0.300 $\pm$ 0.099, p-value=0.020849, 0.1 - the 0.633 $\pm$ 0.047, p-value=0.000264; data were modelled using ANOVA), which we selected for further investigation (Figure 1.

Continuous induction of MCF10A-Plk4 did not enable chronic centrosome amplification

The cell culturing setup consisted of three cell lines, MCF10A-Plk4,MCF10A-Plk4$^{1-608}$, and MCF10A-TetR), each treated with doxycycline at 0, 0.001, 0.1 μg/mL (Figure 1D. We started by growing populations of the three cell lines for 14 days without doxycycline, after which they were sub-sampled and frozen - these were considered to be the ancestral population. We performed two independent experiments using this setup. At the start of an experiment, we thawed one vial for each cell line and seeded cells in 150 mm dishes so they could grow to a large population size (20×10^6 cells). After growing the cells for four days, each population was split into three 150 mm dishes. After 24h, we started the long term experiment by replacing the cell culture medium with medium supplemented with doxycycline at 0, 0.001, or 0.1 μg/mL). We passaged cells every 72h into new 150 mm dishes and added fresh cell culture medium with the appropriate dose of doxycycline. In order to keep a constant bottleneck size, we seeded $2.2x10^6$ cells at every passage. This also ensured that population size did not decrease drastically, thus improving the efficiency of natural selection. Each experiment ran until day 62 after doxycycline addition (20 passages) 1E.

First, we asked how centriole number changed over time. We

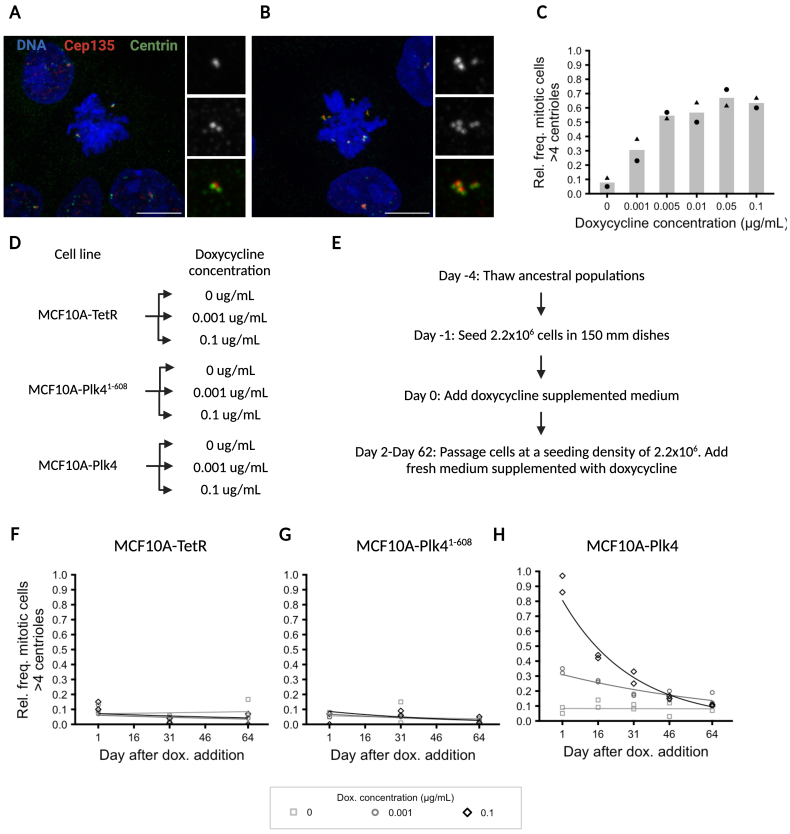


FIGURE 1: **Chronically induced MCF10A-Plk4 recovered wild-type-like centriole numbers.** A-B. Example of mitotic cells with wild-type (A) and abnormally high centriole numbers (B). Cells were stained for Cep135 (red), centrin (green), and DNA (blue). Scale bar is 10 nm. C - Relative frequency of mitotic cells with more than four centrioles after 24h treatment with doxycycline at the indicated concentrations. D - Variables in the long-term cell culturing setup. We grew each cell line (MCF10A-TetR, MCF10A-Plk4$^{1-608}$ at three different concentrations of doxycycline (0, 0.001, and 0.1 μg/mL) for 64 days. E - Long-term cell culture setup. The ancestral population of each cell was thawed and grown for four days, and then split. Doxycycline was added after 24h. Populations were passaged every 72h at a seeding cell number of $2.2x10^6$ for 63 days. F-H - Relative frequency of cells with more than four centrioles at the indicated time points after doxycycline addition for MCF10A-TetR (F), MCF10A-Plk4$^{1-608}$ (G), and MCF10A-Plk4 (H). Data points correspond to populations grown in the presence of doxycycline at 0 (light grey, hollow squares), 0.001 (grey, hollow circles), or 0.1 μg/mL (black, hollow diamonds).

hypothesised that overexpressing Plk4 with higher doses of doxycycline would lead to a chronic increase in the rate of centriole overproduction. Under overproduction-selection balance, we expected that the equilibrium frequency of cells with extra centrioles and/or mean centriole number per cell would increase. To test this hypothesis, we sampled each population at specific time points and scored centriole number mitotic cells as described above. As expected, MCF10A-TetR and MCF10A-Plk4$^{1-608}$ showed little differences in relative frequency of cells with extra centrioles or mean centriole number per cell at days 1, 31, and 64 and independently of doxycycline concentration (Figure 1F and G, S1 and S2). In contrast, the relative frequency of mitotic cells with extra centrioles in induced MCF10A-Plk4 increased initially in a dose-dependent way but decayed exponentially over time. At the final time point, the degree of centrosome amplification in doxycycline-treated MCF10A-Plk4 populations was similar to that of the non-induced populations (Figure 1H). Mean centriole number decreased in similar fashion (Figure S3).

Next, we performed quantitative analysis of these these dynamics. If the processes leading to the disappearance of cells with extra centrioles in MCF10A-Plk4 populations acted identically regardless of doxycycline concentration, we expected that the relative frequency of cells with centrosome amplification decayed at the same rate. As previously noted, for MCF10A-TetR (0 μg/mL - 0.002867 $\pm$ 0.007527, 0.001 μg/mL - -0.008709 $\pm$ 0.012636, 0.1 μg/mL - -0.008101 $\pm$ 0.016310), or MCF10A-Plk4^{1} $-$ 608 (0 μg/mL - -0.007135 $\pm$ 0.015619, 0.001 μg/mL - -0.015172 $\pm$ 0.0040690.1 μg/mL - -0.01841 $\pm$ 0.01077) the results indicated little or no decay. However, for MCF10A-Plk4, the rate of decay increased with doxycycline concentration (0 μg/mL - -0.0004987 ± 0.0069862, 0.001 μg/mL - -0.013021 ± 0.002672, 0.1 μg/mL - -0.034105 ± 0.001985, Figure 1H). Thus, these dynamics indicate different underlying processes in the disappearance

of cells with extra centrioles or that the same processes acted at different magnitudes depending on the dose of doxycycline. In summary, our results indicated that chronic centrosome amplification could not be maintained despite continuous induction of Plk4 overexpression.

The disappearance of cells with extra centrioles was correlated with an increase in relative fitness of induced MCF10A-Plk4 populations

For centriole overproduction-selection balance to hold, both the overproduction rate of centrioles and the strength of negative selection should remain constant. Since induced MCF10A-Plk4 populations seemingly recovered their initial centriole numbers, it is that the selective pressure against cells with extra centrioles increased over time.

To address this hypothesis, we first estimated the intrinsic growth rate of MCF10A-TetR, MCF10A-Plk4$^{1-608}$ and MCF10A-Plk4 relative to the respective population growing without doxycycline. We did not observe significant differences over time in relative growth rate for MCF10A-TetR and MCF10A-Plk4$^{1-608}$ treated with doxycycline compared to populations grown in the absence of doxycycline (Fig 2A and B, MCF10A-TetR: 0.001 μg/mL - 0.02429 ± 0.01763, p-value= 0.171; 0.1 μg/mL - 0.01409 ± 0.01763, p-value=0.426; MCF10A-Plk4$^{1-608}$: 0.001 μg/mL - 0.004093 ± 0.02087, p-value= 0.707; 0.1 μg/mL - -0.003494 ± 0.01087, p-value=0.749,data modelled with a linear model), suggesting minimal effect of doxycycline and/or overexpression of the truncated form Plk4. Conversely, MCF10A-Plk4 experienced a doxycycline-dependent growth defect (Fig. 2C, 0.001 μg/mL - -0.02652 ± 0.02415, p-value= 0.274; 0.1 μg/mL - -0.1181 ± 0.02415, p-value=3.08×10^{-6}) in the first few passages but eventually converged with non-induced populations. The recovery of wild-type-like intrinsic growth rates in induced MCF10A-Plk4 occurred

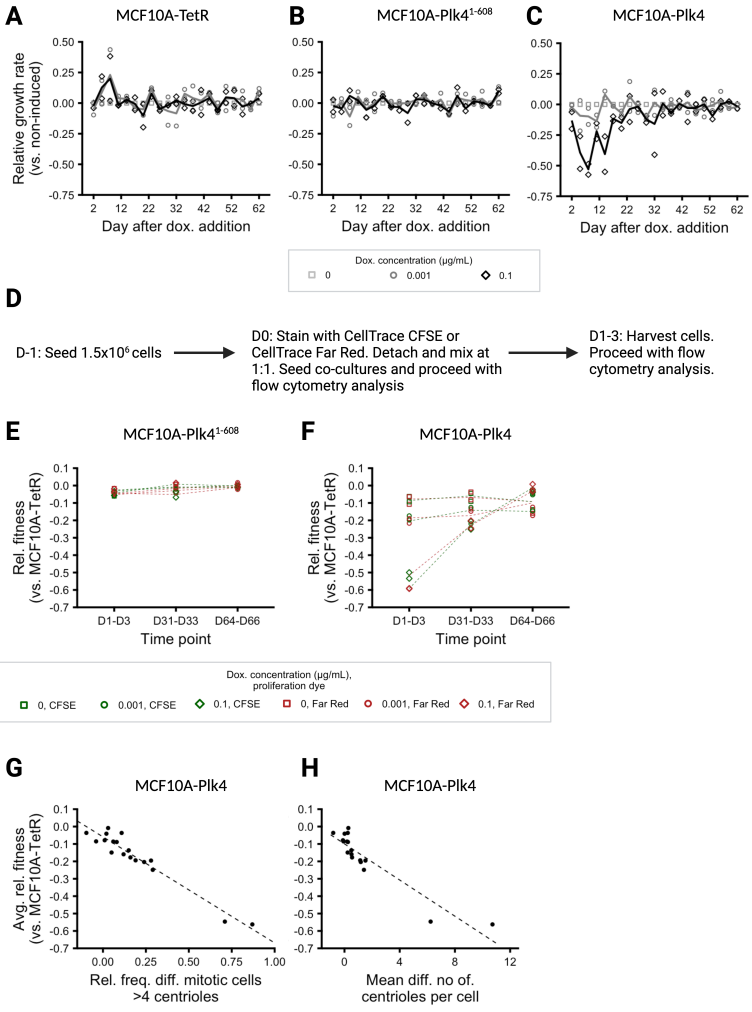

A
MCF10A-TetR
B
MCF10A-Plk4 1-608
C
MCF10A-Plk4
Relative growth rate (vs. non-induced)
0.50
0.25
0.00
-0.25
-0.50
-0.75
2 12 22 32 42 52 62
Day after dox. addition
Dox. concentration (µg/mL)
0
0.001
0.1
D
D-1: Seed 1.5x10^6 cells
D0: Stain with CellTrace CFSE or CellTrace Far Red. Detach and mix at 1:1. Seed co-cultures and proceed with flow cytometry analysis
D1-3: Harvest cells. Proceed with flow cytometry analysis.
E
MCF10A-Plk4 1-608
F
MCF10A-Plk4
Rel. fitness (vs. MCF10A-TetR)
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
D1-D3
D31-D33
D64-D66
Time point
Dox. concentration (µg/mL), proliferation dye
0, CFSE
0.001, CFSE
0.1, CFSE
0, Far Red
0.001, Far Red
0.1, Far Red
G
MCF10A-Plk4
H
MCF10A-Plk4
Avg. rel. fitness (vs. MCF10A-TetR)
0.00 0.25 0.50 0.75 1.00
Rel. freq. diff. mitotic cells >4 centrioles
0 4 8 12
Mean diff. no of. centrioles per cell

FIGURE 2 *(previous page)*: **MCF10A-Plk4 suffered a centriole amplification-dependent fitness cost.** A-C: Growth rate of MCF10A-TetR (A), MCF10A-Plk41 − 608 (B), and MCF10A-Plk4 (C) at each passage of the evolution experiment relative to the non-induced population for each cell line. We estimated relative growth rate by calculating the natural logarithm of the difference between population size at passaging and seeding cell number, and dividing it by the growth rate of the population grown with doxycycline at 0 μg/mL. Data correspond to the population grown with doxycycline at 0 (light gray squares), 0.001 (grey circles), and 0.1 μg/mL (black diamonds). $n_{experiment}$ = 2, $n_{passages}$ = 20 D - Competition assay setup. In brief, we sampled each population at days 0, 29, and 62 after doxycycline addition and seeded 1.5x10^6 cells in T-25 flasks. After 18h, cells were stained with either CellTrace CFSE or CellTrace Far Red, detached, and mixed at 1:1 ratio. Co-cultures were then seeded and harvested after 1-4 days and analysed using flow cytometry. E-F: Relative fitness (competitive index) of MCF10A-Plk4$^{1-608}$ (E) and MCF10A-Plk4 (F) co-cultured with MCF10A-TetR at days 1-3, 31-33 and 64-66 after doxycycline addition. Relative fitness was calculated using the competitive index (see 4). Data points represent co-cultured populations grown with doxycycline at 0 (squares), 0.001 (circles), and 0.1 (diamonds) μg/mL in which MCF10A-Plk4$^{1-608}$ or MCF10A-Plk4 were stained with either CFSE (green, hollow) or Far Red (red, cross). G-H: Correlation between relative fitness averaged across the two proliferation dyes and the difference (G) in relative frequencies of mitotic cells with more than four centrioles between MCF10A-Plk4 and MCF10A-TetR or the difference between mean centriole numbers (H). Relative fitness data include D1-3, D31-33, and D64-66 and centriole number data correspond to D1, D31, and D64. The linear regression is shown as a black line.

concomitantly with the decrease in the relative frequency of cells with extra centrioles.

To gain better insight into the adaptive dynamics of these populations, we performed competitive fitness assays (Fig 2D). In brief, we sampled MCF10A-TetR, MCF10A-Plk4$^{1-608}$, and MCF10A-Plk4 populations at days 0, 30, and 63 after doxycycline addition and grew them for 24h. Afterwards, we stained the cells with either CellTrace CFSE or CellTrace Far Red. CFSE and Far Red are dyes that are stably incorporated by cells, which allows them to be tracked in mono- or co-culture. We then detached and co-cultured each MCF10A-Plk4$^{1-608}$ and MCF10A-Plk4 populations with MCF10A-TetR populations grown in the same conditions. In each pair, one of the populations was stained with CFSE and the other one with Far Red, or vice-versa. Co-cultured populations were then harvested at days 1-3. In addition, we seeded each population stained with either dye in monoculture. As a proxy for competitive fitness, we determined the relative frequency of each CFSE/Far Red single-positive population at each time point and calculated the competitive index (see 4). Essentially, if the relative frequency of one sub-population increased over time we considered it had a competitive fitness advantage over the other, or a disadvantage if it decreased (Figures S4 and S5).

First, we asked if either combination of proliferation dyes affected the outcome of competition. We found a strong correlation in competitive fitness estimates using both dye combinations, suggesting the choice of proliferation dye had little effect in the results for either of cell line pairs (Figure S5A and B; MCF10A-Plk4$^{1-608}$: Pearson's $r \approx 0.55$, p-value≈ 0.018, MCF10A-Plk4: Pearson's $r \approx 0.97$, p-value$\approx 1.18x10^{-11}$). Moreover, we found good correlation between replicates (Figure S5C and D; MCF10A-Plk4$^{1-608}$: Pearson's $r \approx 0.50$, p-value$\approx$0.037, MCF10A-Plk4: Pearson's $r \approx 0.93$, p-value$\approx$ 1.82x10^{-8}) and little effect of the

initial frequency of MCF10A-Plk4$^{1-608}$ or MCF10A-Plk4 on relative fitness estimates (Figure S5E and F; MCF10A-Plk4$^{1-608}$: $\beta_{InitFreq,CFSE}$ = -0.06$\pm$0.13, p-value $\approx$ 0.59; $\beta_{InitFreq,FarRed}$=0.046$\pm$0.10, p-value $\approx$ 0.65;, MCF10A-Plk4: $\beta_{InitFreq,CFSE}$ = -0.27$\pm$0.35, p-value $\approx$ 0.47; $\beta_{InitFreq,FarRed}$=-0.21$\pm$0.27, p-value$\approx$ 0.46).

At days 1-3, 31-33, and 64-66, we inferred a slight intrinsic fitness disadvantage of MCF10A-Plk4 compared to MCF10A-TetR but not MCF10A-Plk4$^{1-608}$. We observed a strong negative effect of the interaction between MCF10A-Plk4 and either dose of doxycycline at days 1-3, which diminished at days 31-33, and was non-significant at days 64-66 (Fig. 2E and F, Table 4.2 and 4.3). These results indicated that induced MCF10A-Plk4 populations suffered an initial doxycycline-dependent fitness disadvantage that improved over time. Finally, we found a strong anti-correlation between competitive fitness and relative frequency of cells with extra centrioles at D1 of the corresponding competition assay (Figure 2G; MCF10A-Plk4$^{1-608}$, Pearson's r $\approx$ 0.97, p-value $\approx 1.47 \times 10^{-11}$). Competitive fitness similarly correlated with mean centriole number but the trend was less linear (Figure 2H). Taken together, doxycycline-treated MCF10A-Plk4 populations suffered a fitness cost proportional to the level of centrosome amplification. However, competitive fitness and intrinsic population growth rates increased over time, suggesting that negative selection against extra centrioles did not play a role in repressing chronic centrosome amplification in the evolved populations.

The fitness cost of extra centrioles depended on slower proliferation

Having inferred a fitness cost associated with extra centrioles, we sought to dissect its underlying causes. In general, we expected cells to be sub-confluent for the duration of the competition assay, such that they should have spent most of the time in exponential

growth. Under exponential growth, fitness differences should depend mostly on cell division and cell death rates.

To quantify the rate of cell division, we took advantage of the chemical properties of the dyes we used to label each co-cultured population. CellTrace CFSE and CellTrace bind covalently to amine groups in cellular proteins. When the cell divides each daughter inherits, on average, half of the dye molecules of its mother. Thus, dye intensity dilutes over time at a rate which is proportional to the rate of cell division. In other words, if the the dye dilutes faster, more cell divisions should have taken place. If it dilutes more slowly, cells should have proliferated less. To determine the rate of dye decay as a proxy for the mean cell division rate, we first calculated the average intensity of CFSE/Far Red in mono- an co-cultured populations at days 1-3 of the competition assay. Then, we performed a linear regression on the logarithm of mean proliferation dye intensity over time (Figures S6 and S7).

We first asked if there was an effect of proliferation dye and replicate on the mean dye decay rate. Our results showed good correlation between the mean dye decay rates obtained with CFSE or Far Red for MCF10A$^{1-608}$ and MCF10A-Plk4 (Figure S7A and B; MCF10A-Plk4$^{1-608}$: Pearson's $r \approx 0.52$, p-value$\approx$ 0.011, MCF10A-Plk4: Pearson's $r \approx 0.69$, p-value$\approx 3.21x10^{-6}$), as well as between replicates (Figure S7C and D; MCF10A-Plk4$^{1-608}$: Pearson's $r \approx 0.48$, p-value$\approx$ 0.028, MCF10A-Plk4: Pearson's $r \approx 0.77$, p-value$\approx 3.32x10^{-8}$). In addition, we asked if there was a difference between mono- and co-cultures and found and either no correlation or good correlation between culture conditions (Figure S7G and H; MCF10A-Plk4$^{1-608}$ (CFSE): Pearson's $r \approx 0.29$, p-value$\approx$ 0.72; MCF10A-Plk4$^{1-608}$ (Far Red): Pearson's $r \approx 0.17$, p-value$\approx$ 0.51, MCF10A-Plk4 (CFSE): Pearson's $r \approx 0.85$, p-value$\approx 6.98x10^{-6}$; MCF10A-Plk4 (Far Red): Pearson's $r \approx 0.75$, p-value$\approx 3.8x10^{-4}$). The lack of correlation between culture

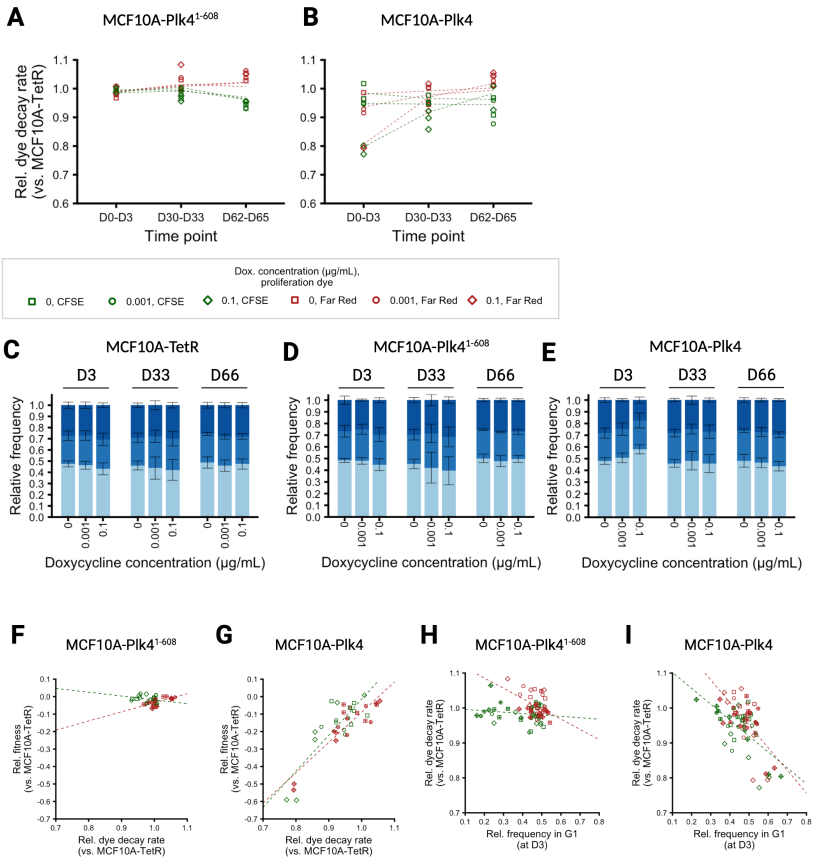


FIGURE 3: **The fitness penalty of induced MCF10A-Plk4 was caused by slower proliferation**. A-B: Relative dye decay rate (proxy for cell division rate) of MCF10A-Plk4$^{1-608}$ (A) and MCF10A-Plk4 stained with CFSE (green) or Far Red (red) and co-cultured with MCF10A-TetR (B) in medium with 0 (squares), 0.001, and 0.1 μg/mL of doxycycline. Relative dye decay rate was estimated by fitting a linear model to the logarithm of the average dye intensity at D1-3 of the competition assay. Data correspond to two independent experiments. C-E - Cell cycle profiles of monocultured MCF10A-TetR (C), MCF10A-Plk4$^{1-608}$ (D), and MCF10A-PLK4 (E) at the indicated time points. The relative frequency of cells in each cell cycle phase (G1 - light blue; S - blue; G2/M - dark blue) was obtained by flow cytometry and manual gating of cells stained with Hoechst 33342. Data correspond to two independent experiments. F-G: Correlation between relative fitness and relative dye decay rate for MCF10A-Plk4$^{1-608}$ (G) and MCF10A-Plk4 (G). Regression lines were obtained by fitting linear models to population labelled with CFSE (green) or Far Red (red). H-I: Correlation between relative fitness and the relative frequency of G1-gated cells for MCF10A-Plk4$^{1-608}$ (G) and MCF10A-Plk4 (H). Regression lines were obtained by fitting linear models to population labelled with CFSE (green) or Far Red (red). Data indicate mono- and co-cultured populations grown with doxycycline at 0 (squares), 0.001 (circles), and 0.1 (diamonds) μg/mL in which MCF10A-Plk4$^{1-608}$ or MCF10A-Plk4 were stained with either CFSE (green, hollow) or Far

conditions for MCF10A-Plk4$^{1-608}$ is probably due to the limited range of variation in mean dye decay rates. Finally, we found no significant effect of the initial frequency of MCF10Aˆ1 − 608 and MCF10A-Plk4 on mean dye decay rate estimates (Figure S7E and F; MCF10A-Plk4$^{1-608}$ (CFSE): $\beta_{InitFreq,CFSE}$=-0.21 ± 0.16, p-value≈ 0.19; MCF10A-Plk4$^{1-608}$ (Far Red): $\beta_{InitFreq,FarRed}$=-0.18 ± 0.13,p-value ≈ 0.19, MCF10A-Plk4 (CFSE): $\beta_{InitFreq,CFSE}$=-0.17 ± 0.14,p-value ≈ 0.24).

The results wd obtained followed a similar trend to those of relative fitness. In quantitative terms, we did not observe an effect of cell line when comparing MCF10A-Plk4 with MCF10A-Plk4$^{1-608}$, or any dose of doxycycline, but we observed an effect of the interaction between MCF10A-Plk4 and both doses of doxycycline at days 1-3 after doxycycline addition; we did not observe an effect at days 31-33 and days 64-66 (Figure 3A and B and Table 4.4 and 4.5). The observed relative fitness differences at days 31-33 and lack thereof in mean dye decay rate are probably to differences in sensitivity in measuring either parameter. Indeed, we observed either no correlation or a positive correlation for MCF10A-Plk4$^{1-608}$ (4F; Pearson's r (CFSE) = 0.28, p-value=0.25; Pearson's r (Far Red)=0.67,p-value=0.0022) but a significant linear correlation between mean dye decay rate and relative fitness for MCF10A-Plk4 3G; Pearson's r (CFSE) = 0.82, p-value=2.61x10^5; Pearson's r (Far Red) = 0.89, p-value=6.88x10^{-7}). These results suggested that the observed fitness cost could be at least partly explained by slower proliferation.

We hypothesised that this proliferation deficit could be due to a delay in the cell cycle. To test this, we stained each population with Hoechst 33342. It allowed us to monitor the state of the cell cycle because its fluorescence intensity increases linearly with DNA concentration. For MCF10A-TetR and MCF10A-Plk4$^{1-608}$ there was little difference in the cell cycle profile depending on doxycycline concentration, dye (CFSE or Far Red) and culture

condition, at days 3, 33, and 66 after doxycycline addition (Figure 3C and D). Conversely, for MCF10A-Plk4, we observed a relative increase in the frequency of G1 cells, with increasing doxycycline concentration at day 3. This trend disappeared at day 33 and day 63 after doxycycline addition (Figure 3E). Finally, we observed that the relative frequency of cells in G1 was not correlated or showed a slight negative correlation with mean dye decay rate for MCF10A-Plk41 − 608 (3F; Pearson's r (CFSE) = 0.28, p-value=0.25; Pearson's r (Far Red)=0.67,p-value=0.0022), but anti-correlated for MCF10A-Plk4 (3G; Pearson's r (CFSE) = -0.70, p-value=2.04x10^{-6}; Pearson's r (Far Red) = -0.54, p-value=0.00077). Taken together, these data indicated that the proliferation disadvantage of induced MCF10A-Plk4 populations was due to cell cycle delay/arrest.

To assess if there were gross differences in cell death rates, we performed an additional staining with propidium iodide (PI). Cells labelled with PI in our competition assay were dead or dying. We found no gross differences in the mean percentage of PI-negative cells across days 1-3 of the competition assay between cell line, doxycycline concentration, CFSE or Far Red staining, or culture conditions at each time point (Figures S8 and S9). Since we did not collect the growth medium of the populations during the competition assay, it is possible that part of the dead cell population that detached from the cell culture plate was lost, such that our assay likely underestimated the effects of cell death. Notwithstanding, our results indicated that a strong effect of extra centrioles on cell death was unlikely.

In conclusion, the fitness deficit induced by extra centrioles was correlated with slower proliferation and cell cycle delay or arrest. The observed increase in both relative intrinsic growth rates and relative fitness over time support adaptation to chronic Plk4 overexpression via the reduction of centrosome amplification in induced MCF10A-Plk4 populations. Ultimately, the data

suggest that the decrease in the relative frequency of cells with extra centrioles in these populations could not be explained by an increasing selective pressure against cells with centrosome amplification.

Plk4 was overexpressed over time but was eventually depleted from the centrosome in induced MCF10A-Plk4 populations

Since the results suggest that induced MCF10A-Plk4 populations adapted to doxycycline addition, we hypothesised that a decrease in the rate of centriole overproduction could explain the decrease in the relative frequency of cells with extra centrioles. Centriole overproduction could be inhibited by transcriptional down-regulation of the exogenous or endogenous Plk4 genes. Alternatively, it is possible that cells lacking the overexpression system could have been present at the beginning of the experiment at low frequencies. To address these possibilities, we quantified relative Plk4 mRNA levels at different time points by quantitative PCR, using specific primers for the exogenous and endogenous transcripts. We observed that the relative mRNA levels of endogenous Plk4 were relatively constant over time in MCF10A-TetR, MCF10A-Plk41 $-$ 608, and MCF10A-Plk4, regardless of doxycycline concentration (Figure 4A). In contrast, the levels of exogenous Plk4 increased with doxycycline concentration at each time point in MCF10A-Plk41 $-$ 608 and MCF10A-Plk4 (MCF10A-Plk41 $-$ 608: 0.001 - 1.302358 $\pm$ 0.332829, p-value$=$ 0.00242; 0.1 - 2.065466 $\pm$ 0.332829, p-value$=$ 6.66×10^{-5}; MCF10A-Plk4: 0.001 - 0.22453 $\pm$ 0.57398, p-value$=$ 0.70387; 0.1 - 2.26653 $\pm$ 0.57398, p-value$=$ 0.00274, linear model with doxycycline concentration and day as predictors) Figures 4B and C). Note that MCF10A-TetR lacks the exogenous transgene. On average, we observed a transient reduction in the abundance of exogenous Plk4 transcripts at days 16 and 31 in MCF10A-Plk4 compared to the other

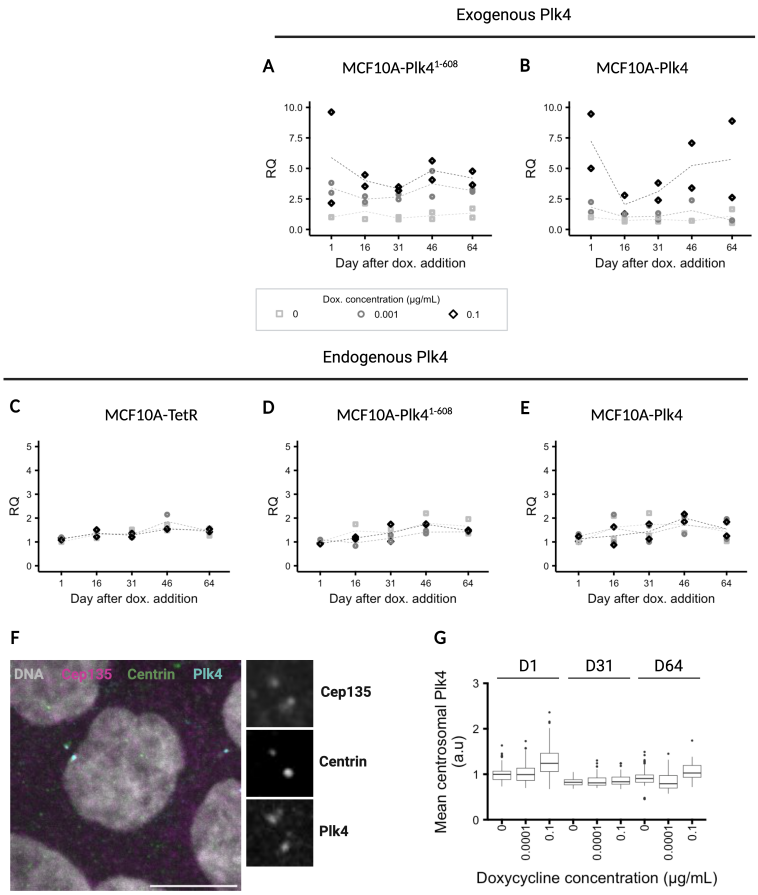


FIGURE 4: **Centrosomal Plk4 levels in evolved MCF10A-Plk4 were reduced despite elevated mRNA levels**. A-B - Exogenous Plk4 mRNA levels in MCF10A-Plk4$^{1-608}$ (A) and MCF10A-Plk4, measured by qPCR (B). Note that MCF10A-TetR lacks the Plk4 transgene. C-E - Endogenous Plk4 mRNA levels in MCF10A-TetR (C), MCF10A-Plk4$^{1-608}$ (D), and MCF10A-Plk4, measured by qPCR. Data refer to populations grown with doxycycline at 0 (squares), 0.001 (circles), and 0.1 (diamonds) μg/mL of the indicated cell line and at the indicated time point. F - Representative image of Plk4 staining. Centrioles were scored based on Cep135 (magenta) and centrin (green) co-localisation. DNA was labelled with Hoechst 33342 (grey). Plk4 foci were identified using automated thresholding and quantified if they co-localised with Cep135/Centrin foci (see Methods). Scale bar is 10 μm. Insets show Cep135 (top), centrin (middle), and Plk4 (bottom) staining. G - Quantification of mean centrosomal Plk4 per cell in MCF10A-Plk4 at D1, D31, and D64. 21-100 cells were scored in one experiment.

.

time points, despite considerable variation within experiments. In MCF10A-Plk41 − 608, this trend was less pronounced and there was less variation between experiments. Nevertheless, exogenous Plk4 was overexpressed in induced MCF10A-Plk4 at days 46 and 64. We observed similar results using a different set of primers specific for exogenous Plk4, as well as primers that recognise both the exogenous and endogenous transcripts (Fig. S10). Therefore, transcriptional silencing of Plk4 or expansion of cells lacking the overexpression system could not explain the steady decrease in the penetrance of centrosome amplification in induced MCF10A-Plk4 populations. Further data is required to assess if the reduction in the levels of Plk4 overexpression at days 16 and 31 could have lowered the rate of centriole overproduction rate at those time points. Next, we asked if the expression of other key centriolar genes could have been affected over time in induced MCF10A-Plk4 populations. As mentioned above, centriole numbers are sensitive to the levels of STIL and SAS-6 (Arquint et al., 2012; Strnad et al., 2007). However, we found no striking differences in the expression of either gene over time, regardless of cell line or doxycycline concentration (Fig. S11.

Having discarded down-regulation of Plk4 mRNA as a possible mechanism for suppressing chronic centrosome amplification, we asked if Plk4 was affected at the protein level. Plk4 is known to localise to the vicinity of mother centrioles, at their proximal end, during the G1-S transition. This event is required to trigger centriole biogenesis (Bettencourt-Dias et al., 2005; Kleylein-Sohn et al., 2007). To test this, we thawed MCF10A-Plk4 cell populations from different time points and stained for Plk4 in order to assess its centrosomal location. Bona fide centrioles and the centrosome were identified by Cep135 and centrin staining as described abova. We observed that Plk4 formed foci in the centrosome that co-localised more closely with Cep135, coherent with its described position at the proximal end of centrioles (Figure

4F). We manually scored these centrosomal foci and obtained a readout of centrosomal Plk4 levels by summing over all such foci in the cell. Due to technical issues concerning the microscope, we could only obtain measurements of Plk4 signal intensity for one of the replicates of the long-term cell culturing setup so the following results are preliminary. In MCF10A-Plk4 populations, the mean intensity of centrosomal Plk4 increased with doxycycline concentration after 24h of induction, as expected (Fig. 4F, 0.001 μg/mL vs. 0 μg/mL: adjusted p-value$\approx$ 1, 0.1 μg/mL vs. 0 μg/mL: adjusted p-value$= 1.398335 \times 10^{-4}$, Dunn's test). However, this effect diminished at day 31 and we observed no significant differences between induced and non-induced populations at day 64 (0.001 μg/mL vs. 0 μg/mL: adjusted p-value$\approx$ 1, 0.1 μg/mL vs. 0 μg/mL: adjusted p-value$= 3.038647 \times 10^{-2}$, Dunn's test) and 64 (0.001 μg/mL vs. 0 μg/mL: adjusted p-value$= 1.315159 \times 10^{-2}$, 0.1 μg/mL vs. 0 μg/mL: adjusted p-value$= 1.604683 \times 10^{-1}$, Dunn's test), consistent with the decrease in the degree of centrosome amplification. These results suggest that post-transcriptional down-regulation of Plk4 or impaired centrosomal recruitment was responsible for preventing chronic centrosome amplification. However, the low sample sizes prevent us from drawing definitive conclusions. In the future, it will be important to repeat this experiment to confirm these results, as well as quantify Plk4 intensity in MCF10A-TetR and MCF10A-Plk4$^{1-608}$ over time as additional controls.

Discussion

With this study, we provided the first systematic assessment of different degrees of chronic perturbation to centriole overproduction-selection balance. Whereas others have reported the effects of chronic Plk4 overexpression at an organismal level (Coelho et al., 2015; Kulukian et al., 2015; Levine et al., 2017; Vitre et al., 2015),

centriole number dynamics and the temporal response to different levels of expression had been sparsely described. To the best of our knowledge, this work is the first example where several cell physiological parameters relative to the level of centrosome amplification were accurately tracked over an extensive period of time. We found that cell populations continuously overexpressing a Plk4 transgene at different levels returned to wild-type-like centriole numbers. Moreover, the rate at which cells carrying extra centrioles disappeared from the population increased with the exposure to higher levels of Plk4. Our results indicated that fitness of cell populations decreased as the relative frequency of cells with extra centrioles rose. This fitness cost could be explained by decreased proliferation due to a cell cycle delay/arrest. We observed no overt increase in cell death in induced MCF10A-Plk4. In addition, we observed an increase in relative fitness as extra centrioles disappeared. At the end of the long-term cell culture experiment, Plk4 mRNA levels remained elevated in a doxycycline-dependent way. However, we observed a reduction in levels of centrosomal Plk4. Overall, our data support a role for negative selection in limiting the expansion of cells carrying extra centrioles, but chronic centrosome amplification under chronic Plk4 overexpression was likely abolished by post-transcriptional regulation of Plk4 levels/activity or compromised centrosomal recruitment.

Considerations of this study

The number of variables combinatorial design, and duration of the experimental evolution setup, as the maintenance of a large population size in order to maximise the effect of selection, deemed that each experiment was a significant investment of time and resources, which limited our ability to scale up this project and perform additional replicates. However, both experiments yielded qualitatively identical and quantitatively similar

results. Moreover, we observed identical trends for different doses of doxycycline, which further strengthen our conclusions. It should also be emphasised that each long-term experiment was carried out independently. Since we did not evolve replicate populations, our results should not be interpreted in the context of parallel evolution, and we cannot draw any conclusions about the causes of specific evolutionary trajectories.

Concerning the generality of our results, it is pertinent to note that different tissue types show widely variable responses to the presence of extra centrioles (Coelho et al., 2015; Kulukian et al., 2015; Levine et al., 2017; Serçin et al., 2016; Shoshani et al., 2021; Vitre et al., 2015). Although there are few examples that can be compared to this study, Vitre *et al.* (Vitre et al., 2015) generated a cell line where Plk4 can be constitutively overexpressed following Ad Cre infection. The authors observed that the relative frequency of cells with extra centrioles stabilised and persisted around 40% for one month. The same study reports another inducible cell line in which the level of centrosome amplification decreased after induction of Plk4 overexpression. Along these lines, Levine *et al.* (Levine et al., 2017) observed continuously decreasing levels of centrosome amplification upon chronic doxycycline-mediated induction of Plk4 overexpression in mouse embryionic fibroblasts. Indeed, the relative frequency of interphase cells with extra centrioles decreased over 14 days. These results indicate that different cell types may respond differently to chronic Plk4 overexpression. It will be important to repeat these experiments using different model systems and experimental conditions to better understand in which contexts centriole overproduction-selection balance can be stably altered or not. In particular, it will be important to quantify centriole number dynamics upon continuous overexpression of other key centriole biogenesis regulators, such as STIL or SAS-6.

Regarding the competition assays, there are some technical aspects that should be considered when interpreting their results. Since MCF10A-Plk4 is an adherent cell line, we seeded replicate populations to be harvested at each day of the assay rather than sampling the same population at different time points. Moreover, for technical reasons related to viability/cell cycle staining and sample size during data acquisition, we seeded each replicate population such that they reached the same approximate confluence at the time of harvest. Thus, the seeding density was adjusted accordingly for each day of harvest. It is possible part of the effect on the estimated relative fitness and dye decay rate (as a proxy for the cell division) may be due to the different seeding densities. Nevertheless, relative fitness values did not differ greatly when using data from different time points, suggesting our results are robust to this factor. Moreover, relative fitness estimates generally agreed with the intrinsic growth rate estimates obtained from population sizes at passaging (where the seeding density is kept constant). Notwithstanding, it should be noted that the underlying fitness and proliferation dynamics in the competition assay be more complex. For this reason, we also refrained from estimating population doubling times based on proliferation dye dilution.

Implications for centriole number dynamics and regulation of Plk4 abundance

Previous reports suggested that chronic centrosome amplification could not be maintained *in vivo* or *in vitro* under chronic Plk4 overexpression in the presence of an intact p53 pathway (Vitre et al., 2015). However, a modest level of Plk4 overexpression might facilitate chronic centrosome amplification *in vivo* but possibly not *in vitro* (Levine et al., 2017). By systematically comparing modest and severe Plk4 overexpression scenarios, our data showed that MCF10A-Plk4 populations reverted to basal

levels of centrosome amplification in either case. Therefore, *in vitro*, low levels of Plk4 overexpression did not allow for sustained centrosome amplification.

Furthermore, our results seemingly contradict the general predictions of centriole overproduction-selection balance. This hypothesis posits that the distribution of centriole numbers per cell will tend towards a stable equilibrium if the rate of centriole overproduction and the selective pressure of cells with extra centrioles balance each other out (Dias Louro et al., 2021a). According to our results, it is likely that induced MCF10A-Plk4 experienced a decrease in the centriole overproduction rate over time, which drove the population to the original equilibrium. We were unable to pinpoint the source of this decrease but it may depend transiently on transcriptional silencing of the exogenous Plk4 promoter, as suggested by the reduction in mRNA levels at days 16 and 31, and some form of post-transcriptional regulation which leads to the degradation of Plk4, inhibits its kinase activity, or impairs its centrosomal recruitment. For instance, Plk4 can target itself for degradation via the SCF-Slimb E3 ubiquitin ligase complex (Cunha-Ferreira et al., 2013; Cunha-Ferreira et al., 2009; Guderian et al., 2010) and may also regulate the abundances of other key centriolar proteins (Puklowski et al., 2011) but, theoretically, either of these mechanisms could be possible. Such a phenotype could have been established via a negative feedback loop triggered by Plk4 overexpression or centrosome amplification, or it might have been present in the ancestral population, since it is polyclonal, having risen in frequency due to being favoured by selection. The fact that relative fitness estimates predict a faster decrease in the level of centrosome amplification than what was observed support the idea that the decrease in the relative frequency of cells with extra centrioles was caused by a physiological response rather than "resistant" cells rising in frequency. Irrespective of this, our results indicate that the

doxycycline-treated populations suffered a decrease in the centriole overproduction rate, thus preventing centriole numbers from attaining a different equilibrium. Quantification of Plk4 levels at the cytosol may help narrowing down where regulation may be occurring and mRNA sequencing at different could allow us to discern if there were changes in the expression levels of known regulators of Plk4 stability.

These mechanisms can potentially explain the difference in dynamics between modest and severe Plk4 overexpression. If the reduction in the relative frequency of cells with extra centrioles was due to a negative feedback loop triggered by Plk4 overexpression or centrosome amplification itself, then a higher frequency of cells carrying extra centrioles should more efficiently inhibit further centriole overproduction in the population. Likewise, if the ancestral population consisted of cells capable of generating extra centrioles and cells insensitive to centrosome amplification, then the latter should sweep through more rapidly if a higher fraction of the former do produce excess centrioles. This can be justified by the observation that relative fitness decreases linearly with the degree of centrosome amplification. In the future, it will be important to tease apart these two mechanisms. Mathematical modelling may offer a straightforward approach to test which one best explains the data.

Another important aspect to consider for the dynamics of chronic centrosome amplification is p53 activity. In cultured cell lines, there is broad evidence that p53 is stabilised upon Plk4 overexpression (Fava et al., 2017; Holland et al., 2012; Levine et al., 2017), including in MCF10A-Plk4 (Godinho et al., 2014). Knocking out or inhibiting p53 is known to rescue the associated proliferation defects and can allow extra centrioles to be maintained or, at least, delay their disappearance. This is also true in some tissues *in vivo* (Coelho et al., 2015; Kulukian et al., 2015; Vitre et al., 2015). It has also been suggested that p53 may regulate Plk4 expression

directly or indirectly (Holland et al., 2012; Vitre et al., 2015. Thus, p53 activation could, in principle, account for both the initial proliferation deficit in doxycycline-treated populations and for suppression of centriole overproduction. While we expect p53 to be stabilised as an initial response to Plk4 overexpression, it was probably not be responsible for inhibiting of centrosome amplification over longer time scales because population growth rates returned to basal levels and p53 activation should lead to cell cycle arrest (Fava et al., 2017; Holland et al., 2012). Nevertheless, it will be important to tease apart the relationship of p53 and extra centrioles or Plk4 overexpression in doxycycline-treated populations and test if and when it was active during the experiments.

Perspectives on cancer evolution

The causes of centrosome amplification in live tumours remain obscure. Whereas Plk4, STIL, and SAS-6, as well as other genes associated with centriole numbers are commonly found be over-expressed in cancer (Arquint and Nigg, 2016), most mechanisms responsible for regulating their abundances operate at the protein level (Arquint et al., 2012; Cunha-Ferreira et al., 2013; Cunha-Ferreira et al., 2009; Guderian et al., 2010; Puklowski et al., 2011; Strnad et al., 2007). Thus, their expression levels may not necessarily be predictive of centrosome amplification. Our results support this idea because centriole overproduction was abrogated despite abnormally high Plk4 expression levels.

Regardless of their source, extra centrioles must arise and persist in the population in order to have a role in cancer evolution. The strength of negative selection against cells with centrosome amplification should be a critical parameter in determining whether or not this should occur. If selection is strong, any cell carrying extra centrioles should be efficiently eliminated. Although we estimated a heavy fitness penalty associated with extra centrioles,

cell populations were able to avoid extinction even under severe Plk4 overexpression and in the absence of further alterations which could allow them to cope with centrosome amplification.

In this regard, it is also important to consider how selection relates to number of extra centrioles or to penetrance of cells with centrosome amplification. It was previously proposed that low number of extra centrioles may be pro-tumourigenic whereas high numbers are too deleterious and may suppress cancer development (Levine et al., 2017). Indeed, mice overexpressing Plk4 at modest levels developed tumours containing a small fraction of cells with low numbers of extra centrioles. However, *in vitro* mouse embryonic fibroblasts carrying the same construct suffered a significant proliferation deficit. Despite bearing low numbers of extra centrioles, cells with centrosome amplification reached high frequencies in these populations. In agreement, our results suggest that the relative frequency of cells with extra centrioles was a better predictor of fitness than mean centriole number. In other words, our data indicates that populations containing few cells with centrosome amplification are fitter than populations where cells with centrosome amplification are more prevalent.

The relevance of this result lies in the fact that centriole overproduction tends to occur stochastically (Dias Louro et al., 2021b; Lopes et al., 2015). For example, inducing Plk4 overexpression in an isogenic population can yield cells with a random number of extra centrioles or even cells without centrosome amplification. Suppose that a mutation leading to a slight increase in the rate of centriole overproduction arises at some point in a pre-malignant tissue. For an individual cell, such a mutation might not even be costly if it does not induce errors in centriole biogenesis. For the whole population, the fitness cost said mutation could impart might be negligible. Neutral processes could then drive the mutation to fixation. Thus, genetic or phenotypic variation

leading to slight increases in the error rate of centriole biogenesis can theoretically be maintained in the population. In coherence, cells in live tumours generally contain few extra centrioles (Chan, 2011; Godinho and Pellman, 2014; Lopes et al., 2018).

If this is true, it poses the question of how extra centrioles can affect cancer evolution if they are relatively rare or unstable in the population. One possibility is through aneuploidy. Centrosome amplification can lead to chromosome segregation errors. A single one of these errors can lead to amplification of an oncogene or to loss of a tumour suppressor (Nicholson and Cimini, 2015). In addition, it has been shown that transient centrosome amplification is sufficient for certain cancer types to arise (Levine et al., 2017; Shoshani et al., 2021). Finally, centrosome amplification can promote invasion via paracrine signalling (Arnandis et al., 2018). In this case, centrosome amplification might even be beneficial, at least for neighbouring cells. Thus, however unlikely it may be, it is plausible that tumourigenesis or increased malignancy could be triggered by a few number of cells containing extra centrioles.

Nevertheless, this possibility is still a matter of speculation. In the future it will be important to accurately quantify cell survival and ploidy with respect to the number of extra centrioles. It will also be crucial to test if a population of cells with low levels of centrosome amplification can be maintained when competing with a "normal" population, although finding an adequate system for testing this hypothesis might be challenging. Irrespective of this, the quantitative relationship we have established between relative fitness and penetrance of centrosome amplification allows us to predict that if centrosome amplification is present in the early stages of tumourigenesis, before any coping mechanisms are established, it should be relatively rare. This has been observed, for instance, in Barrett's oesophagus, a condition which precedes the appearance of gastroesophageal adenocarcinoma (Lopes et al., 2018).

Methods

Cell culture

We cultured MCF10A-TetR, MCF10A-Plk4$^{1-608}$, and MCF10A-Plk4 as described previously by Godinho *et al.* (Godinho et al., 2014). In short, we grew the cells in DMEM (Gibco) and F-12 (Gibco) at 1:1 ratio, supplemented with horse serum at, L-glutamine (Thermo Scientific) at 0.5 μg mL^{-1}, penicillin and streptomycin (Thermo Scientific) at 100 U mL^{-1} 100 insulin at 10 μg mL^{-1} (Sigma), hydrocortisone (Sigma) at 100 ng mL^{-1}, cholera toxin (Sigma) at 1 μ g mL^{-1}, epidermal growth factor (Sigma) at 20 ng mL^{-1}, and maintained them at 37° C in a 5% CO_2 atmosphere. All cell lines were tested for the presence of mycoplasma.

For the evolution experiment, we passaged cells every 72h and subcultured them at a density of 2.2x10^6 cells in 150 mm dishes containing 20 mL of complete cell culture medium. This seeding density ensured that cells were sub-confluent at the time of passaging, such that they predominantly spent time in exponential growth. The medium for doxcycyline-treated populations was further supplemented with doxycycline (Sigma) at 0.001 or 0.1 μg mL^{-1}. Cell culture medium was replaced every 72h in the evolving populations, during passaging. During passaging, we diluted the cell suspension at 1:90 and performed automated cell counting using a Scepter 2.0 Handheld Automated Cell Counter (Merck Millipore) and 60 μm Sensors (Merck Millipore).

The ancestral populations for each cell line were generated by growing them in complete cell culture medium for 14 days in 150 mm dishes. We froze the cells in cryotubes with complete cell culture medium containing 10% of DMSO and stored them overnight at -80° C in Mr. Frosty freezing containers (Thermo Scientific) before transferring them to liquid nitrogen tanks. During each long-term experiment, we froze cells for each condition

at D11, D26, D41, D56, and D62 using medium with the appropriate doxycycline concentration and according to the protocol described above. When thawing cells, we seeded them in 150 mm dishes filled with the appropriate cell culture medium, which was replaced 24h after thawing.

Competition assays and flow cytometry

We seeded approximately $1.5x10^6$ cells cells at D-1, D29, and D62 of each experimental evolution replicate in two T-25 flasks per condition, such that confluence was not reached in the next day and cells maintained exponential growth. At day 0 of each competition assay, CellTrace CFSE (Thermo Scientific) or CellTrace Far Red (Thermo Scientific) was performed according to manufacturer instructions. In brief, we replaced the cell culture medium in the flasks with 10 (CellTrace CFSE) and 2 μM (CellTrace Far Red) in 3 mL of 1 x PBS and incubated the cells at 37° C in 5% CO_2 for 20 minutes. After staining, we washed the flasks with an equal volume of medium supplemented with 5% horse serum to remove unbound dye molecules. We detached the cells and incubated them at 37° C in 5% CO_2 for 15 minutes. Then, cells were ressuspended and transferred to 2 mL tubes. Co-cultures were prepared by mixing populations stained with CellTrace CFSE and CellTrace Far Red at 1:1 volume ratio. We seeded monocultures and co-cultures in 6-well microplates at 1:7.5 dilution from the cell suspension for day 1, 1:18 for day 2, 1:45 for day 3 such that the confluence at the time of harvest was approximately 70-80%. Before proceeding with flow cytometry analysis, we stained cells with propidium iodide (PI, Sigma) at 1 μg mL^{-1} and incubated them at 37° C. At days 1 through to 3, cells were stained with Hoechst 33342 (Invitrogen) at 5 μg mL^{-1} for 30 minutes at 37° C in 5% CO_2 before detaching them and performing PI staining as described above. Mono- and co-cultures were then analysed by flow cytometry. At day 0 of the

competition assay we analysed the co-cultures by flow cytometry to assess the relative frequencies of each sub-population.

Flow cytometry analysis was conducted using a BD Fortessa X-20 cytometer and BD FACSDiva software. Data acquisition was performed using a low flow rate. Voltages and gating strategy for flow cytometry analysis were optimised using single-staining and fluorescence-minus-one controls independently for day 0 and days 1-3 samples. The cell population was gated based on their forward- and side-scatter area parameters (FSC-A and SSC-A, respectively) and doublets were excluded by plotting the area against the width of the forward-scatter parameter. Exclusion of dead cells was performed by gating out PI-positive cells. Finally, we gated live single-positive CellTrace CFSE and CellTrace FarRed cells for competition and proliferation analysis. Cell cycle analysis was conducted in Hoechst 33342-positive and PI-negative cells. We acquired at least 20,000 events for each sample. Flow cytometry data was processed and visualised using R.

Quantification of relative fitness and dye decay rate

As mentioned in the previous section, we first gated single-positive CellTrace CFSE and CellTrace Far Red populations. We observed a small population of double-positive cells, as reported in the literature, which we excluded from further analysis. The relative frequency, p, of each sub-population labelled with dye i, at day t was calculated according to the following expressions:

$$p_i(t) = \frac{N_i(t)}{N_i(t) + N_i(t)} \tag{4.1}$$

where $N_i(t)$ is the total number of single-positive cells for CellTrace CFSE or CellTrace Far Red. As a proxy for fitness, we employed the competitive index, given by:

$$w = ln\Big(\frac{RelFreq^{final}_{Comp}/RelFreq^{final}_{Comp}}{RelFreq^{init}_{Comp}/RelFreq^{init}_{Ref}}\Big) \tag{4.2}$$

To calculate the mean rate of dye decay, as a proxy for the mean cell division rate, we first computed the average CFSE/Far Red intensity in single-positive cells at each day of the competition assay. As above, double-positive cells were excluded. As a simplification, we assumed constant dye decay and estimated the mean decay rate by performing a linear regression on the logarithm of the differences of average dye intensity.

Estimation of relative fitness and mean dye decay rate was performed using R.

Immunofluorescence and microscopy

For staining and imaging of centriolar proteins, cells were grown on 13 mm coverslips for 24h (samples from experimental evolution) or 48h (for additional experiments) in 24-well plates, using the appropriate cell culture medium. Afterwards, we aspirated the medium and fixed cells in 100% methanol at -20ºC for 10′. Fixed samples were stored in 1xPBS (REF) at 4ºC until use. For immunostaining, we started by incubating cells with blocking solution (10% FBS in 1xPBS) for 30′-1h at room temperature. Primary antibodies were dissolved in blocking solution and samples were incubated overnight at 4ºC. Secondary antibody solutions were prepared in the same way as primary antibody solutions and spun at 15,000 rpm for 5′ to remove aggregates. We washed samples in 1xPBS and incubated them for 1h at room temperature on secondary antibody solutions. Then, samples were once again washed with 1xPBS and DNA was stained with Hoechst 33342 at 1 μg/mL. Finally, we mounted coverslips on microscopy slides, using Vectashield mounting medium(Vector Laboratories). Primary antibodies used: anti-Cep135 C-terminus 1:500 (rabbit,

Abcam), anti-Cep135 1:250 (rat, Metabion, raised against human CEP135 a.a. 271-371 expressed in *E. coli* as a fusion protein with a N-terminal 10X His-tag.), anti-centrin clone 20H5 1:500 (mouse, Millipore); anti-hPlk4 1:250 (rabbit, Metabion, produced against the c-terminal region of human Plk4 a.a. 510-970 and purified using the membrane-bound c-terminal region of Plk4). Secondary antibodies were conjugated with: Alexa Fluor 488 1:500 (raised in goat/donkey, anti-mouse, Molecular Probes), DyLight 488 1:250 (REFERENCE 1:250), Rhodamine Red 1:500 (raised in donkey, anti-rabbit, Jackson), Alexa 647 (raised in donkey, anti-rat).

Images of mitotic cells for assessing centriole numbers and interphase cells for quantification of Plk4 levels were acquired using an Eclipse Ti-2 inverted microscope with 3i Marianas spinning-disk using a 100X objective. We scored 100 mitotic/interphasic cells except where noted. Image analysis was performed using Fiji/ImageJ. Centrioles number counting was performed on 3D stacks. Each centrin focus co-localising with a Cep135 focus was scored as a centriole.

Quantification of centrosomal Plk4 levels by immunofluorescence

To quantify Plk4 fluorescence intensity we started by performing maximum intensity projections of the Z-stack and applying an automatic threshold to the corresponding channel using the built-in Renyi Entropy algorithm in Fiji/ImageJ. ROIs were then manually selected if they co-localised with the centriolar markers centrin and CEP135. Raw integrated intensity of each selected ROI was measured in the sum intensity projection. To measure the background, we moved each ROI to an adjacent, non-overlapping, position and quantified the aforementioned statistic. The Plk4 signal for each ROI was calculated by taking its raw integrated density and subtracting the corresponding background values. These values were summed over all the

ROIs in each cell to calculate the total raw integrated density of the centrosomal Plk4 signal. Finally, we measured the mean (corrected for the background) and area of the centrosomal Plk4 signal per ROI and for each cell.

RNA extraction

Cells were grown for 24h in 6-well plates in the appropriate cell culture medium, after which we detached and pelleted them. After aspirating the supernatant, cells were frozen in dry ice and stored at -80°C until use. RNA extraction was performed immediately after thawing the pellets using the RNeasy Mini Kit (Qiagen) according to manufacturer instructions. For each sample, we performed on-column DNA digestion using RNase-free DNAse (Qiagen) as specified in the kit. RNA quantity and purity were assessed using NanoDrop 2000 (Thermo Scientific).

qPCR

We performed cDNA synthesis using the High Capacity RNA-to-cDNA kit (Applied Biosystems), according to manufacturer instructions. For the first long-term cell culturing experiment, we synthesised cDNA from 1.5 μg of RNA. For the second experiment, we used 0.5 μg of RNA because one of the samples was less concentrated.

For the qPCR reaction we used iTaq SYBR Green (BioRad) according to kit instructions. Samples were prepared in triplicate in 384-well plates. The qPCR run and data analysis were performed using QuantStudio (Applied Biosystems). The following primers were used for quantifying the mRNA levels of GAPDH (endogenous control), exogenous truncated or full-length Plk4, endogenous Plk4, total Plk4 (i.e. primers anneal to both the exogenous and endogenous sequences), STIL, and SAS-6.

TABLE 4.1: List of primers

Target	Sequence	Orientation	Reference
GAPDH	5′-TTAAAAGCAGCCCTGGTGAC-3′	Forward	Godinho et al., 2014
GAPDH	5′-CTCTGCTCCTCCTGTTCGAC-3′	Reverse	Godinho et al., 2014
Exogenous Plk4	5′-CAGGATTTGCCCGGGATGGCG-3′	Forward	Godinho et al., 2014
Exogenous Plk4	5′-AACCAGTGTGAATGGACTCAGCTCT-3′	Reverse	Godinho et al., 2014
Exogenous Plk4 #2	5′-TTTCCGAGGAGGATTTGCCC-3′	Forward	This study
Exogenous Plk4 #2	5′-ACCAGTGTGAATGGACTCAGC-3′	Reverse	This study
Endogenous Plk4	5′-CTAATCCGGAGAACCCAGGC-3′	Forward	This study
Endogenous Plk4	5′-ACCAGTGTGAATGGACTCAGC-3′	Reverse	This study
Total Plk4	5′-AGACCACCCTTCGACACTGA-3′	Forward	This study
Total Plk4	5′-GTCCTTGGCCTCTATTGACAAA-3′	Reverse	This study
STIL	5′-AATGAAGTCACAAGTCTTCCAGG-3′	Forward	This study
STIL	5′-CACAACTAGAGAAGAGCTGTTGG-3′	Reverse	This study
SAS-6	5′-GAATGGGCGTCACATACAGC-3′	Forward	This study
SAS-6	5′-TTGATATTGAACCTGTGCCTGC-3′	Reverse	This study

Data visualisation and statistical analyses

Plots and statistical analyses were produced using R. Figures were created using BioRender (https://app.biorender.com/).

Supplementary figures

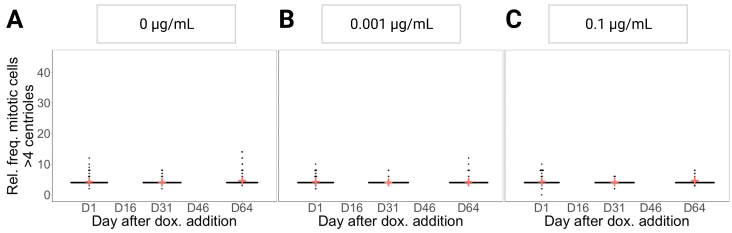


FIGURE S1: **Centriole number distributions for MCF10A-TetR**. Data represent the number of centrioles per cell in MCF10A-TetR grown with doxycycline at 0 (A), 0.001 (B), or 0.1 (C) μg/mL at days 1, 31, and 64 after doxycycline addition. The mean is indicated in red. $n_{experiment} = 2$.

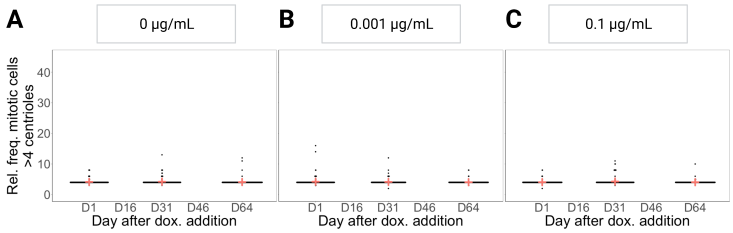


FIGURE S2: **Centriole number distributions for MCF10A-Plk4$^{1-608}$**. Data represent the number of centrioles per cell in MCF10A-Plk4$^{1-608}$ grown with doxycycline at 0 (A), 0.001 (B), or 0.1 (C) μg/mL at days 1, 31, and 64 after doxycycline addition. The mean is indicated in red. $n_{experiment} = 2$.

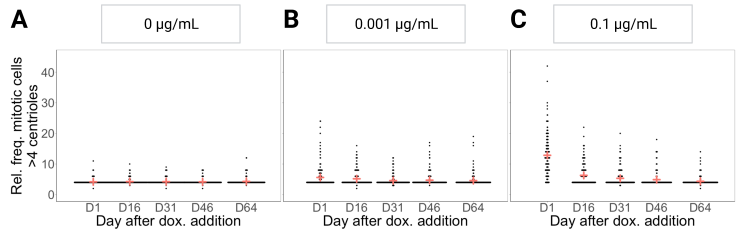


FIGURE S3: **Centriole number distributions for MCF10A-Plk4**. Data represent the number of centrioles per cell in MCF10A-Plk4 grown with doxycycline at 0 (A), 0.001 (B), or 0.1 (C) μg/mL at days 1, 31, and 64 after doxycycline addition. The mean is indicated in red. $n_{experiment} = 2$.

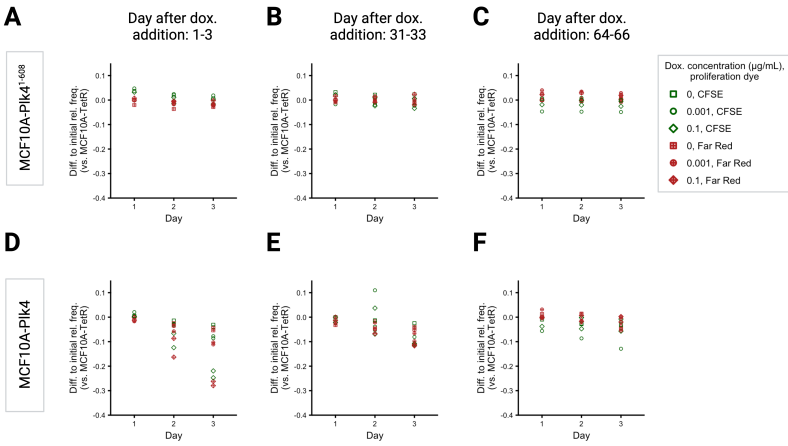


FIGURE S4: **Temporal variation in the relative frequencies of MCF10A-Plk4$^{1-608}$ or MCF10A-Plk4 co-culture with MCF10A-TetR**. Difference to the initial frequency of MCF10A-Plk4$^{1-608}$ (A-C) and MCF10A-Plk4 (D-F) at days 1-3 (A,D), 31-33 (B,E), and 62-65 (C,F) after doxycycline addition. Data points indicate co-cultured populations grown with doxycycline at 0 (squares), 0.001 (circles), and 0.1 (diamonds) μg/mL in which MCF10A-Plk4$^{1-608}$ or MCF10A-Plk4 were stained with either CFSE (green, hollow) or Far Red (red, cross). $n_{experiment} = 2$.

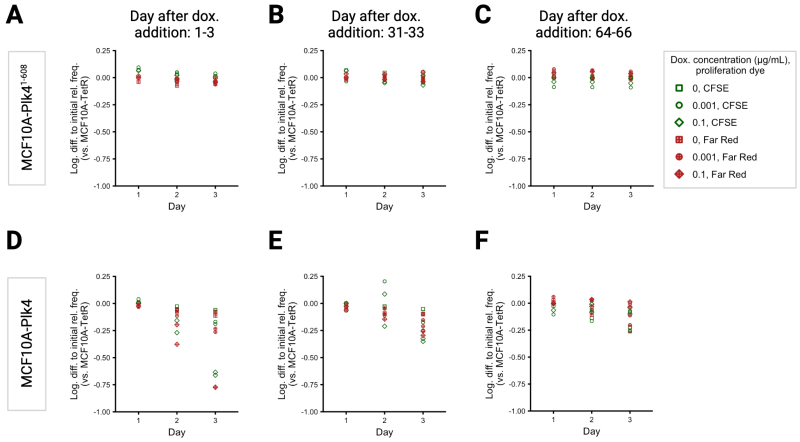


FIGURE S5: **Temporal variation in the natural logarithm of the relative frequencies of MCF10A-Plk4$^{1-608}$ or MCF10A-Plk4 co-culture with MCF10A-TetR**. Logarithm of the difference to the initial frequency of MCF10A-Plk4$^{1-608}$ (A-C) and MCF10A-Plk4 (D-F) at days 1-3 (A,D), 31-33 (B,E), and 62-65 (C,F) after doxycycline addition. Data points indicate co-cultured populations grown with doxycycline at 0 (squares), 0.001 (circles), and 0.1 (diamonds) μg/mL in which MCF10A-Plk4$^{1-608}$ or MCF10A-Plk4 were stained with either CFSE (green, hollow) or Far Red (red, cross). $n_{experiment} = 2$.

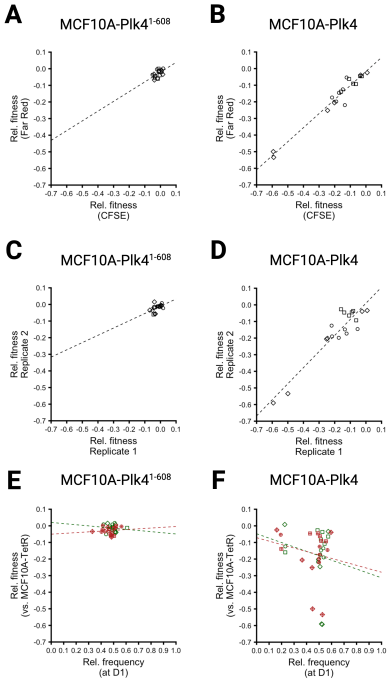

A
MCF10A-Plk4 1-608
Rel. fitness (Far Red)
Rel. fitness (CFSE)
B
MCF10A-Plk4
Rel. fitness (Far Red)
Rel. fitness (CFSE)
C
MCF10A-Plk4 1-608
Rel. fitness Replicate 2
Rel. fitness Replicate 1
D
MCF10A-Plk4
Rel. fitness Replicate 2
Rel. fitness Replicate 1
E
MCF10A-Plk4 1-608
Rel. fitness (vs. MCF10A-TetR)
Rel. frequency (at D1)
F
MCF10A-Plk4
Rel. fitness (vs. MCF10A-TetR)
Rel. frequency (at D1)

FIGURE S5 *(previous page)*: **Effects of proliferation dye, experimental replicates, initial frequencies, and culture conditions on relative fitness estimates.** A-B: Correlation between proliferation dyes in estimated relative fitness for MCF10A-Plk4$^{1-608}$ (A) and MCF10A-Plk4 (B) populations. The data include and populations co-cultured with MCF10A-TetR at D1-3, D31-33, and D64-66 from two independent experiments. Linear regression is shown as a dashed line and indicates X correlation between relative fitness estimated in CFSE- versus Far Red-stained MCF10A-Plk41 − 608 populations and Y correlation for MCF10A-Plk4. C-D: Correlation between each independent experiment in estimated relative dye decay rates for MCF10A-Plk4$^{1-608}$ (C) and MCF10A-Plk4 (D). The data include populations co-cultured with MCF10A-TetR stained with either CFSE or Far Red at D1-3, D31-33, and D64-66. Linear regression is shown as a dashed line and indicates X correlation between replicates in the relative dye decay rates estimated in CFSE- and Far Red-stained MCF10A-Plk41 − 608 populations and Y correlation for MCF10A-Plk4. E-F: Correlation between estimated relative dye decay rates and initial frequency for MCF10A-Plk4$^{1-608}$ (E) and MCF10A-Plk4 (F). The data include populations co-cultured with MCF10A-TetR stained with either CFSE or Far Red at D1-3, D31-33, and D64-66 in two-independent experiments. Linear regression is shown as a dashed line and indicates X correlation between replicates in the relative dye decay rates estimated in CFSE- and Far Red-stained MCF10A-Plk41 − 608 populations and Y correlation for MCF10A-Plk4. Data points indicate co-cultured populations grown with doxycycline at 0 (squares), 0.001 (circles), and 0.1 (diamonds) μg/mL in which MCF10A-Plk4$^{1-608}$ or MCF10A-Plk4 were stained with either CFSE (green) or Far Red (red).

[!htb]

TABLE 4.2: **Predictors for relative fitness in CFSE-labelled populations.**

	Day 1-3	Day 31-33	Day 64-66
β_0	-0.026182 ± 0.012362, p-value$= 0.079$	-0.007723 ± 0.017154, p-value$= 0.66834$	-0.012102 ± 0.031352, p-value$= 0.713$
$\beta_{MCF10A-Plk4}$	-0.059854 ± 0.017482, p-value$= 0.01408$	-0.051591 ± 0.024259, p-value$= 0.07756$	-0.080895 ± 0.044339, p-value$= 0.118$
$\beta_{Dox0.001}$	-0.017779 ± 0.017482, p-value$= 0.34840$	0.016184 ± 0.024259, p-value$= 0.52949$	0.009548 ± 0.044339, p-value$= 0.837$
$\beta_{Dox0.1}$	-0.004408 ± 0.017482, p-value$= 0.80935$	-0.004103 ± 0.024259, p-value$= 0.87125$	0.012384 ± 0.044339, p-value$= 0.789$
$\beta_{MCF10A-Plk4*Dox0.001}$	-0.099619 ± 0.024724, p-value$= 0.00689$	-0.098325 ± 0.034307, p-value$= 0.02858$	-0.064920 ± 0.062705, p-value$= 0.340$
$\beta_{MCF10A-Plk4*Dox0.1}$	-0.099619 ± 0.024724, p-value$= 0.00689$	-0.098325 ± 0.034307, p-value$= 0.02858$	-0.064920 ± 0.062705, p-value$= 0.340$

[!htb]

TABLE 4.3: **Predictors for relative fitness in Far Red-labelled populations.**

	Day 1-3	Day 31-33	Day 64-66
β_0	-0.052296 ± 0.012939, p-value$= 0.00679$	-0.02871 ± 0.02489, p-value$= 0.2926$	-0.006715 ± 0.027033, p-value$= 0.8121$
$\beta_{MCF10A-Plk4}$	-0.025474 ± 0.018298, p-value$= 0.21329$	-0.03911 ± 0.03520, p-value$= 0.3090$	-0.085689 ± 0.038230, p-value$= 0.0662$
$\beta_{Dox0.001}$	0.005507 ± 0.018298, p-value$= 0.77359$	0.01235 ± 0.03520, p-value$= 0.7377$	0.005466 ± 0.038230, p-value$= 0.8910$
$\beta_{Dox0.1}$	0.007572 ± 0.018298, p-value$= 0.69340$	-0.02230 ± 0.03520, p-value$= 0.5498$	-0.004112 ± 0.038230, p-value$= 0.9179$
$\beta_{MCF10A-Plk4*Dox0.001}$	-0.113823 ± 0.025878, p-value$= 0.00458$	-0.11693 ± 0.04978, p-value$= 0.0571$	-0.012631 ± 0.054065, p-value$= 0.8230$
$\beta_{MCF10A-Plk4*Dox0.1}$	-0.446710 ± 0.025878, p-value$= 2.42 \times 10^{-}6$	-0.13862 ± 0.04978, p-value$= 0.0318$	-0.064636 ± 0.054065, p-value$= 0.2770$

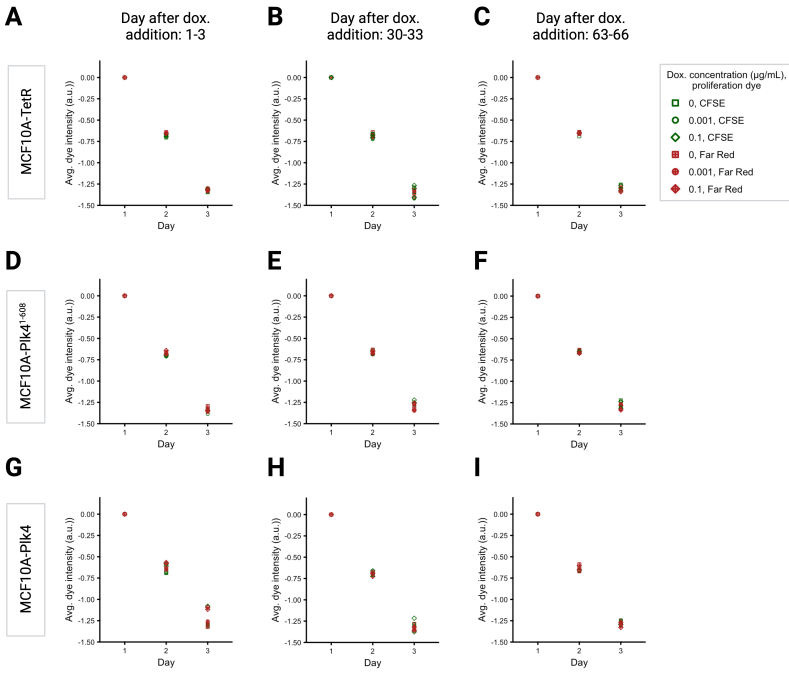


FIGURE S6: **Temporal variation in average dye intensity of mono-cultured cell lines**. Difference from the initial average dye intensity of MCF10A-TetR (A-C) MCF10A-Plk4$^{1-608}$ (D-F) and MCF10A-Plk4 (G-I) at days 1-3 (A, D, G), 31-33 (B, E, H), and 62-65 (C, F, I) after doxycycline addition. Data indicate co-cultured populations grown with doxycycline at 0 (squares), 0.001 (circles), and 0.1 (diamonds) μg/mL in which MCF10A-Plk4$^{1-608}$ or MCF10A-Plk4 were stained with either CFSE (green, hollow) or Far Red (red, cross). Note that the y-axis is in log-scale. $n_{experiment} = 2$.

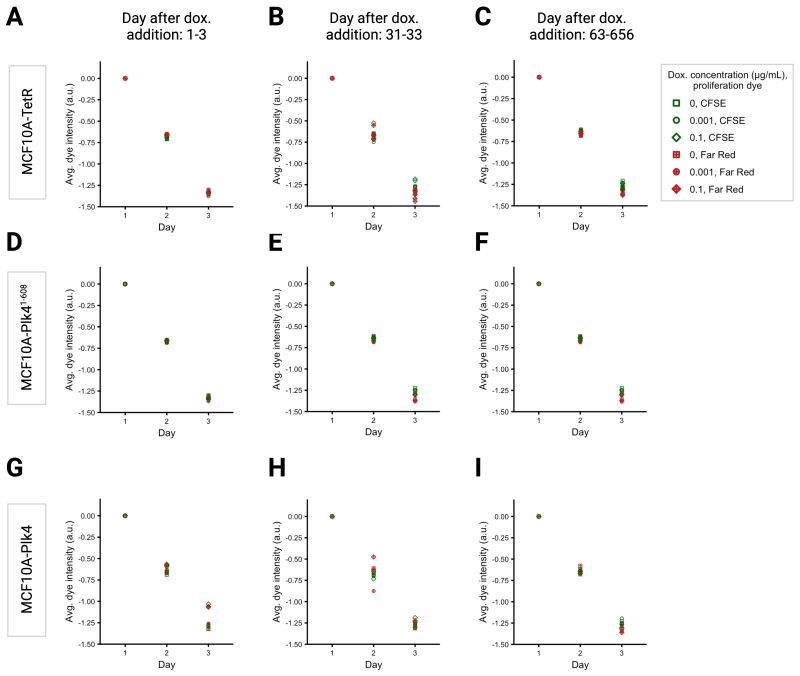


FIGURE S7: **Temporal variation in average dye intensity of co-cultured cell lines**. Difference from the initial average dye intensity of MCF10A-TetR (A-C) MCF10A-Plk4$^{1-608}$ (D-F) and MCF10A-Plk4 (G-I) at days 1-3 (A, D, G), 31-33 (B, E, H), and 64-66 (C, F, I) after doxycycline addition. Data indicate co-cultured populations grown with doxycycline at 0 (squares), 0.001 (circles), and 0.1 (diamonds) μg/mL in which MCF10A-Plk4$^{1-608}$ or MCF10A-Plk4 were stained with either CFSE (green, hollow) or Far Red (red, cross). Note that the y-axis is in log-scale. The data for MCF10A-TetR include co-cultures with both MCF10A-Plk4$^{1-608}$ and MCF10A-Plk4. $n_{experiment} = 2$.

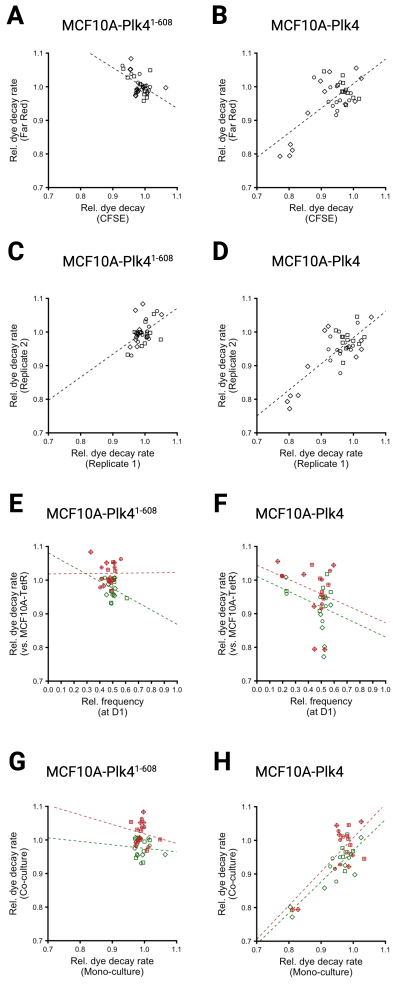

A
MCF10A-Plk4 1-608
Rel. dye decay rate (Far Red)
Rel. dye decay (CFSE)
B
MCF10A-Plk4
Rel. dye decay rate (Far Red)
Rel. dye decay (CFSE)
C
MCF10A-Plk4 1-608
Rel. dye decay rate (Replicate 2)
Rel. dye decay rate (Replicate 1)
D
MCF10A-Plk4
Rel. dye decay rate (Replicate 2)
Rel. dye decay rate (Replicate 1)
E
MCF10A-Plk4 1-608
Rel. dye decay rate (vs. MCF10A-TetR)
Rel. frequency (at D1)
F
MCF10A-Plk4
Rel. dye decay rate (vs. MCF10A-TetR)
Rel. frequency (at D1)
G
MCF10A-Plk4 1-608
Rel. dye decay rate (Co-culture)
Rel. dye decay rate (Mono-culture)
H
MCF10A-Plk4
Rel. dye decay rate (Co-culture)
Rel. dye decay rate (Mono-culture)

FIGURE S7 *(previous page)*: **Effects of proliferation dye, experimental replicates, initial frequencies, and culture conditions on relative dye decay rate estimates**. A-B: Correlation between proliferation dyes in estimated relative dye decay rates for MCF10A-Plk4$^{1-608}$ (A) and MCF10A-Plk4 (B) populations. The data include mono-cultured populations and populations co-cultured with MCF10A-TetR at D1-3, D31-33, and D64-66 from two independent experiments. Linear regression is shown as a dashed line and indicates X correlation between relative dye decay rates estimated in CFSE- versus Far Red-stained MCF10A-Plk41 − 608 populations and Y correlation for MCF10A-Plk4. C-D: Correlation between each independent experiment in estimated relative dye decay rates for MCF10A-Plk4$^{1-608}$ (C) and MCF10A-Plk4 (D). The data include mono-cultured populations and populations co-cultured with MCF10A-TetR stained with either CFSE or Far Red at D1-3, D31-33, and D64-66. Linear regression is shown as a dashed line and indicates X correlation between replicates in the relative dye decay rates estimated in CFSE- and Far Red-stained MCF10A-Plk41 − 608 populations and Y correlation for MCF10A-Plk4. E-F: Correlation between estimated relative dye decay rates and initial frequency for MCF10A-Plk4$^{1-608}$ (E) and MCF10A-Plk4 (F). The data include populations co-cultured with MCF10A-TetR stained with either CFSE or Far Red at D1-3, D31-33, and D64-66 in two-independent experiments. Linear regression is shown as a dashed line and indicates correlation between replicates in the relative dye decay rates estimated in CFSE- and Far Red-stained MCF10A-Plk41 − 608 populations and Y correlation for MCF10A-Plk4. G-H: Correlation between estimated relative dye decay rates in mono-cultured and co-cultured MCF10A-Plk4$^{1-608}$ (G) and MCF10A-Plk4 (H). The data include populations stained with either CFSE or Far Red at D1-3, D31-33, and D64-66 in two independent experiments. Linear regression is shown as a dashed line and indicates X correlation between culture conditions in the relative dye decay rates estimated in CFSE- and Far Red-stained MCF10A-Plk41 − 608 populations and Y correlation for MCF10A-Plk4. Data points indicate co-cultured populations grown with doxycycline at 0 (squares), 0.001 (circles), and 0.1 (diamonds) μg/mL in which MCF10A-Plk4$^{1-608}$ or MCF10A-Plk4 were stained with either CFSE (green) or Far Red (red).

[!htb]

TABLE 4.4: **Predictors for relative dye dilution rate in co-cultured CFSE-labelled populations.**

	Day 1-3	Day 31-33	Day 64-66
β_0	1.000533 ± 0.013280 p-value$= 3.68 \times 10{-}10$	0.98267 ± 0.01417 p-value$= 6.06 \times 10^{-10}$	0.939349 ± 0.027372 p-value$= 4.08 \times 10^{-8}$
$\beta_{MCF10A-Plk4}$	-0.009631 ± 0.018781, p-value$= 0.626405$	-0.01853 ± 0.02005, p-value$= 0.3910$	-0.001263 ± 0.038710, p-value$= 0.975$
$\beta_{Dox0.001}$	0.003869 ± 0.018781, p-value$= 0.843611$	0.01117 ± 0.02005 p-value$= 0.5977$	0.004545 ± 0.038710 p-value$= 0.910$
$\beta_{Dox0.1}$	-0.003242 ± 0.018781, p-value$= 0.868639$	-0.01671 ± 0.02005, p-value$= 0.4364$	0.015258 ± 0.038710, p-value$= 0.707$
$\beta_{MCF10A-Plk4*Dox0.001}$	-0.046838 ± 0.026561, p-value$= 0.128289$	-0.03992 ± 0.02835, p-value$= 0.2087$	-0.024326 ± 0.054744, p-value$= 0.672$
$\beta_{MCF10A-Plk4*Dox0.1}$	-0.200918 ± 0.026561, p-value$= 0.000277$	-0.0695 ± 0.02835, p-value$= 0.0497$	0.013695 ± 0.054744, p-value$= 0.811$

[!htb]

TABLE 4.5: **Predictors for relative dye dilution rate in co-cultured Far Red-labelled populations.**

	Day 1-3	Day 31-33	Day 64-66
β_0	0.983901 ± 0.010267, p-value$= 8.7 \times 10^{-11}$	1.016121 ± 0.032201, p-value$= 6.73 \times 10^{-8}$	1.040462 ± 0.010841, p-value$= 8.62 \times -11$
$\beta_{MCF10A-Plk4}$	-0.008670 ± 0.014519, p-value$= 0.5722$	-0.041243 ± 0.045540 p-value$= 0.400$	-0.012047 ± 0.015331, p-value$= 0.462$
$\beta_{Dox0.001}$	0.002522 ± 0.014519, p-value$= 0.8678$	0.004796 ± 0.045540, p-value$= 0.920$	0.009638 ± 0.015331, p-value$= 0.553$
$\beta_{Dox0.1}$	0.010997 ± 0.014519, p-value$= 0.4775$	0.022240 ± 0.045540, p-value$= 0.643$	0.011076 ± 0.015331, p-value$= 0.497$
$\beta_{MCF10A-Plk4*Dox0.001}$	-0.055934 ± 0.020533, p-value$= 0.0345$	-0.001272 ± 0.064403, p-value$= 0.985$	-0.017672 ± 0.021682, p-value$= 0.446$
$\beta_{MCF10A-Plk4*Dox0.1}$	-0.192342 ± 0.020533, p-value$= 8.4 \times 10^{-5}$	-0.027830 ± 0.064403, p-value$= 0.681$	0.010429 ± 0.021682, p-value$= 0.648$

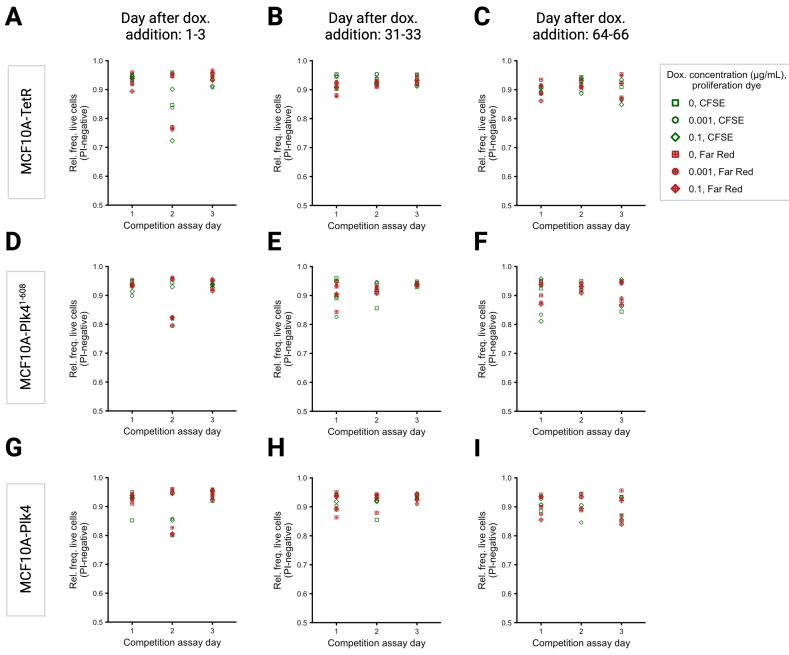


FIGURE S8: **Doxycycline treatment did not induce differential cell death in mono-cultures.** Cells were stained with propidium iodide (PI), which can enter dead or dying cells with compromised cell membranes, to assess viability. A-C: Relative frequency of live (PI-negative) MCF10A-TetR. D-F: Relative frequency of live MCF10A-Plk4$^{1-608}$. G-I Relative frequency of live MCF10A-Plk4 (G-I). Shown are the data corresponding to cells stained at days 1-3 (A, D, G), 31-33 (B, E, H), and 64-66 (C, F, I) after doxycycline addition during each competition assay. Data points indicate mono-cultured populations grown with doxycycline at 0 (squares), 0.001 (circles), and 0.1 (diamonds) μg/mL in which MCF10A-Plk4$^{1-608}$ or MCF10A-Plk4 were stained with either CFSE (green) or Far Red (red). $n_{experiment} = 2$.

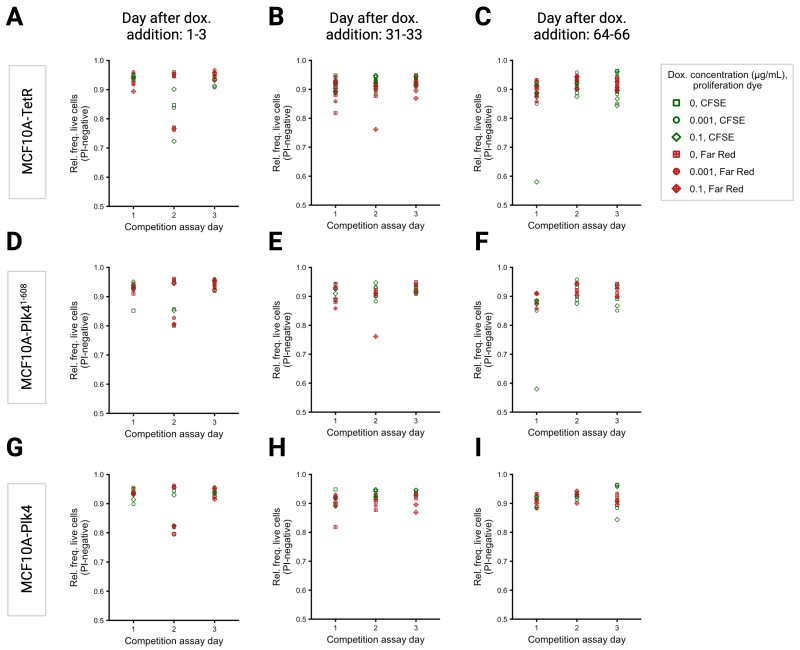


FIGURE S9: **Doxycycline treatment did not induce differential cell death in co-cultures**. Cells were stained with propidium iodide (PI), which can enter dead or dying cells with compromised cell membranes, to assess viability. A-C: Relative frequency of live (PI-negative) MCF10A-TetR. D-F: Relative frequency of live MCF10A-Plk4$^{1-608}$. G-I Relative frequency of live MCF10A-Plk4 (G-I). Shown are the data corresponding to cells stained at days 1-3 (A, D, G), 31-33 (B, E, H), and 64-66 (C, F, I) after doxycycline addition during each competition assay. Data points indicate co-cultured populations grown with doxycycline at 0 (squares), 0.001 (circles), and 0.1 (diamonds) μg/mL in which MCF10A-Plk4$^{1-608}$ or MCF10A-Plk4 were stained with either CFSE (green) or Far Red (red). The data for MCF10A-TetR include co-cultures with both MCF10A-Plk4 and MCF10A-Plk4$^{1-608}$.

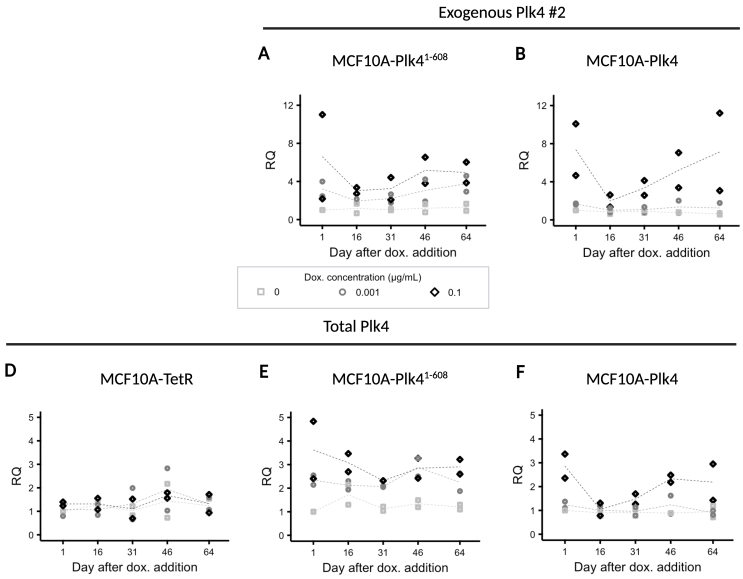


FIGURE S10: **Plk4 was overexpressed in MCF10A-Plk4 and MCF10A-Plk4$^{1-608}$ after two months of doxycycline treatment**. Relative quantification of mRNA levels of the indicated genes at different time points, by qPCR. GAPDH was used as an endogenous control and the non-induced condition at day 1 after doxycycline addition was used as a reference. A-B: Relative mRNA levels of exogenous Plk4, using a different set of primers, in MCF10A-Plk4$^{1-608}$ (A), and MCF10A-Plk4 (B). C-E: Relative mRNA levels of total Plk4, using a pair of primers that target both the endogenous and exogenous sequences, in MCF10A-TetR (C), MCF10A-Plk4$^{1-608}$ (D), and MCF10A-Plk4 (E). Data points correspond to populations grown in the presence of doxycycline at 0 (light grey, hollow squares), 0.001 (grey, hollow circles), or 0.1 μg/mL (black, hollow diamonds).

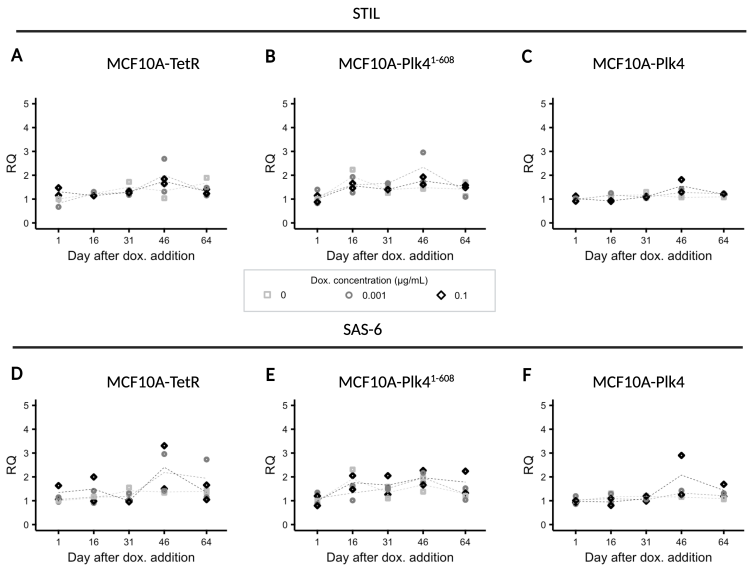


FIGURE S11: **The expression levels of STIL and SAS-6 showed little variation over time.** Relative quantification of mRNA levels of the indicated genes at different time points, by qPCR. GAPDH was used as an endogenous control and the non-induced condition at day 1 after doxycycline addition was used as a reference. A-C: Relative mRNA levels of STIL in MCF10A-TetR (A), MCF10A-Plk4$^{1-608}$ (B), and MCF10A-Plk4 (C). D-F: Relative mRNA levels of SAS-6 in MCF10A-TetR (A), MCF10A-Plk4$^{1-608}$ (B), and MCF10A-Plk4 (C). Data points correspond to populations grown in the presence of doxycycline at 0 (light grey, hollow squares), 0.001 (grey, hollow circles), or 0.1 μg/mL (black, hollow diamonds).

References

Arnandis, T., Monteiro, P., Adams, S. D., Bridgeman, V. L., Rajeeve, V., Gadaleta, E., Marzec, J., Chelala, C., Malanchi, I., Cutillas, P. R., & Godinho, S. A. (2018). Oxidative Stress in Cells with Extra Centrosomes Drives Non-Cell-Autonomous Invasion. *Developmental Cell*, *47*(4), 409–424.e9. https://doi.org/10.1016/j.devcel.2018.10.026

Arquint, C., Sonnen, K. F., Stierhof, Y.-D., & Nigg, E. A. (2012). Cell-cycle-regulated expression of STIL controls centriole number in human cells. *Journal of Cell Science*, *125*(5), 1342–1352. https://doi.org/10.1242/jcs.099887

Arquint, C., & Nigg, E. A. (2016). The PLK4–STIL–SAS-6 module at the core of centriole duplication. *Biochemical Society Transactions*, *44*(5), 1253–1263. https://doi.org/10.1042/bst20160116

Baudoin, N. C., Nicholson, J. M., Soto, K., Martin, O., Chen, J., & Cimini, D. (2020). Asymmetric clustering of centrosomes defines the early evolution of tetraploid cells. *eLife*, *9*. https://doi.org/10.7554/eLife.54565

Bettencourt-Dias, M., Rodrigues-Martins, A., Carpenter, L., Riparbelli, M., Lehmann, L., Gatt, M. K., Carmo, N., Balloux, F., Callaini, G., & Glover, D. M. (2005). SAK/PLK4 is required for centriole duplication and flagella development. *Current Biology*, *15*(24), 2199–2207. https://doi.org/10.1016/j.cub.2005.11.042

Bettencourt-Dias, M., Hildebrandt, F., Pellman, D., Woods, G., & Godinho, S. A. (2011). Centrosomes and cilia in human disease. *Trends in Genetics*, *27*(8), 307–315. https://doi.org/10.1016/j.tig.2011.05.004

Breslow, D. K., & Holland, A. J. (2019). Mechanism and Regulation of Centriole and Cilium Biogenesis. *Annual Review of Biochemistry*, *88*(1), 691–724. https://doi.org/10.1146/annurev-biochem-013118-111153

Brito, D. A., Gouveia, S. M., & Bettencourt-Dias, M. (2012). Deconstructing the centriole: Structure and number control. *Current Opinion in Cell Biology*, *24*(1), 4–13. https://doi.org/10.1016/j.ceb.2012.01.003

Chan, J. Y. (2011). A clinical overview of centrosome amplification in human cancers. *Int. J. Biol. Sci.*, *7*(8), 1122–1144.

Coelho, P. A., Bury, L., Shahbazi, M. N., Liakath-Ali, K., Tate, P. H., Wormald, S., Hindley, C. J., Huch, M., Archer, J., Skarnes, W. C., Zernicka-Goetz, M., & Glover, D. M. (2015). Over-expression of Plk4 induces centrosome amplification, loss of primary cilia and associated tissue hyperplasia in the mouse. *Open Biology*, *5*(12). https://doi.org/10.1098/rsob.150209

Cunha-Ferreira, I., Bento, I., Pimenta-Marques, A., Jana, S. C., Lince-Faria, M., Duarte, P., Borrego-Pinto, J., Gilberto, S., Amado, T., Brito, D., Rodrigues-Martins, A., Debski, J., Dzhindzhev, N., & Bettencourt-Dias, M. (2013). Regulation of autophosphorylation controls PLK4 self-destruction and centriole number. *Current Biology*, *23*(22), 2245–2254. https://doi.org/10.1016/j.cub.2013.09.037

Cunha-Ferreira, I., Rodrigues-Martins, A., Bento, I., Riparbelli, M., Zhang, W., Laue, E., Callaini, G., Glover, D. M., & Bettencourt-Dias, M. (2009). The SCF/Slimb Ubiquitin Ligase Limits Centrosome Amplification through Degradation of SAK/PLK4. *Current Biology*, *19*(1), 43–49. https://doi.org/10.1016/j.cub.2008.11.037

Dias Louro, M. A., Bettencourt-Dias, M., & Bank, C. (2021a). Patterns of selection against centrosome amplification in human cell lines. *PLOS Computational Biology*, *17*(5), 1–22. https://doi.org/10.1371/journal.pcbi.1008765

Dias Louro, M. A., Bettencourt-Dias, M., & Carneiro, J. (2021b). A first-takes-all model of centriole copy number control

based on cartwheel elongation. *PLOS Computational Biology*, *17*(5), 1–22. https://doi.org/10.1371/journal.pcbi.1008359

Fava, L. L., Schuler, F., Sladky, V., Haschka, M. D., Soratroi, C., Eiterer, L., Demetz, E., Weiss, G., Geley, S., Nigg, E. A., & Villunger, A. (2017). The PIDDosome activates p53 in response to supernumerary centrosomes. *Genes and Development*, *31*(1), 34–45. https://doi.org/10.1101/gad.289728.116

Galofré, C., Asensio, E., Ubach, M., Torres, I. M., Quintanilla, I., Castells, A., & Camps, J. (2020). Centrosome reduction in newly-generated tetraploid cancer cells obtained by separase depletion. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-65975-1

Godinho, S. A., & Pellman, D. (2014). Causes and consequences of centrosome abnormalities in cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1650). https://doi.org/10.1098/rstb.2013.0467

Godinho, S. A., Picone, R., Burute, M., Dagher, R., Su, Y., Leung, C. T., Polyak, K., Brugge, J. S., Théry, M., & Pellman, D. (2014). Oncogene-like induction of cellular invasion from centrosome amplification. *Nature*, *510*(7503), 167–171. https://doi.org/10.1038/nature13277

Gomes Pereira, S., Dias Louro, M. A., & Bettencourt-Dias, M. (2021). Biophysical and Quantitative Principles of Centrosome Biogenesis and Structure. *Nature Communications*, *37*, 43–63. https://doi.org/10.1038/s41467-017-00238-8.

Gönczy, P. (2015). Centrosomes and cancer: Revisiting a long-standing relationship. *Nature Reviews Cancer*, *15*(11), 639–652. https://doi.org/10.1038/nrc3995

Guderian, G., Westendorf, J., Uldschmid, A., & Nigg, E. A. (2010). Plk4 trans-autophosphorylation regulates centriole number by controlling trcp-mediated degradation. *Journal of*

Cell Science, *123*(13), 2163–2169. https://doi.org/10.1242/jcs.068502

Holland, A. J., Fachinetti, D., Zhu, Q., Bauer, M., Verma, I. M., Nigg, E. A., & Cleveland, D. W. (2012). The autoregulated instability of polo-like kinase 4 limits centrosome duplication to once per cell cycle. *Genes Dev.*, *26*(24), 2684–2689.

Karsenti, E. (2008). Self-organization in cell biology: A brief history. https://doi.org/10.1038/nrm2357

Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Habedanck, R., Stierhof, Y.-D., & Nigg, E. A. (2007). Plk4-induced centriole biogenesis in human cells. *Developmental Cell*, *13*(2), 190–202. https://doi.org/https://doi.org/10.1016/j.devcel.2007.07.002

Kulukian, A., Holland, A. J., Vitrec, B., Naika, S., Cleveland, D. W., & Fuchsa, E. (2015). Epidermal development, growth control, and homeostasis in the face of centrosome amplification. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(46), E6311–E6320. https://doi.org/10.1073/pnas.1518376112

Kwon, M., Godinho, S. A., Chandhok, N. S., Ganem, N. J., Azioune, A., Thery, M., & Pellman, D. (2008). Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes. *Genes and Development*, *22*(16), 2189–2203. https://doi.org/10.1101/gad.1700908

Lambrus, B. G., Uetake, Y., Clutario, K. M., Daggubati, V., Snyder, M., Sluder, G., & Holland, A. J. (2015). P53 protects against genome instability following centriole duplication failure. *Journal of Cell Biology*, *210*(1), 63–77. https://doi.org/10.1083/jcb.201502089

Levine, M. S., Bakker, B., Boeckx, B., Moyett, J., Lu, J., Vitre, B., Spierings, D. C., Lansdorp, P. M., Cleveland, D. W., Lambrechts, D., Foijer, F., & Holland, A. J. (2017). Centrosome

Amplification Is Sufficient to Promote Spontaneous Tumorigenesis in Mammals. *Developmental Cell*, *40*(3), 313–322.e5. https://doi.org/10.1016/j.devcel.2016.12.022

Lopes, C. A. M., Jana, S. C., Cunha-Ferreira, I., Zitouni, S., Bento, I., Duarte, P., Gilberto, S., Freixo, F., Guerrero, A., Francia, M., Lince-Faria, M., Carneiro, J., & Bettencourt-Dias, M. (2015). PLK4 trans-Autoactivation Controls Centriole Biogenesis in Space. *Developmental Cell*, *35*(2), 222–235. https://doi.org/10.1016/j.devcel.2015.09.020

Lopes, C. A., Mesquita, M., Cunha, A. I., Cardoso, J., Carapeta, S., Laranjeira, C., Pinto, A. E., Pereira-Leal, J. B., Dias-Pereira, A., Bettencourt-Dias, M., & Chaves, P. (2018). Centrosome amplification arises before neoplasia and increases upon p53 loss in tumorigenesis. *Journal of Cell Biology*, *217*(7), 2353–2363. https://doi.org/10.1083/jcb.201711191

Nicholson, J. M., & Cimini, D. (2015). Link between aneuploidy and chromosome instability. *International Review of Cell and Molecular Biology*, *315*, 299–317. https://doi.org/10.1016/bs.ircmb.2014.11.002

Ohta, M., Ashikawa, T., Nozaki, Y., Kozuka-Hata, H., Goto, H., Inagaki, M., Oyama, M., & Kitagawa, D. (2014). Direct interaction of Plk4 with STIL ensures formation of a single procentriole per parental centriole. *Nature communications*, *5*, 5267. https://doi.org/10.1038/ncomms6267

Ohta, M., Watanabe, K., Ashikawa, T., Nozaki, Y., Yoshiba, S., Kimura, A., & Kitagawa, D. (2018). Bimodal Binding of STIL to Plk4 Controls Proper Centriole Copy Number. *Cell Reports*, *23*(11), 3160–3169.e4. https://doi.org/10.1016/j.celrep.2018.05.030

Puklowski, A., Homsi, Y., Keller, D., May, M., Chauhan, S., Kossatz, U., Grünwald, V., Kubicka, S., Pich, A., Manns, M. P., Hoffmann, I., Gönczy, P., & Malek, N. P. (2011). The SCF-FBXW5 E3-ubiquitin ligase is regulated by PLK4 and

targets HsSAS-6 to control centrosome duplication. *Nature Cell Biology*, *13*(8), 1004–1009. https://doi.org/10.1038/ncb2282

Sala, R., Farrell, K., & Stearns, T. (2020). Growth disadvantage associated with centrosome amplification drives population-level centriole number homeostasis. *Molecular Biology of the Cell*, mbc.E19–04–0195. https://doi.org/10.1091/mbc.e19-04-0195

Serçin, Ö., Larsimont, J. C., Karambelas, A. E., Marthiens, V., Moers, V., Boeckx, B., Le Mercier, M., Lambrechts, D., Basto, R., & Blanpain, C. (2016). Transient PLK4 overexpression accelerates tumorigenesis in p53-deficient epidermis. *Nature Cell Biology*, *18*(1), 100–110. https://doi.org/10.1038/ncb3270

Shoshani, O., Bakker, B., de Haan, L., Tijhuis, A. E., Wang, Y., Kim, D. H., Maldonado, M., Demarest, M. A., Artates, J., Zhengyu, O., Mark, A., Wardenaar, R., Sasik, R., Spierings, D. C. J., Vitre, B., Fisch, K., Foijer, F., & Cleveland, D. W. (2021). Transient genomic instability drives tumorigenesis through accelerated clonal evolution. *Genes Dev.*, *35*(15-16), 1093–1108.

Strnad, P., Leidel, S., Vinogradova, T., Euteneuer, U., Khodjakov, A., & Gönczy, P. (2007). Regulated HsSAS-6 Levels Ensure Formation of a Single Procentriole per Centriole during the Centrosome Duplication Cycle. *Developmental Cell*, *13*(2), 203–213. https://doi.org/10.1016/j.devcel.2007.07.004

Takao, D., Watanabe, K., Kuroki, K., & Kitagawa, D. (2019a). Feedback loops in the Plk4–STIL–HsSAS6 network coordinate site selection for procentriole formation. *Biology Open*, *8*(9). https://doi.org/10.1242/bio.047175

Takao, D., Yamamoto, S., & Kitagawa, D. (2019b). A theory of centriole duplication based on self-organized spatial pattern formation. *Journal of Cell Biology*, *218*(11), 3537–3547. https://doi.org/10.1083/jcb.201904156

Vitre, B., Holland, A. J., Kulukian, A., Shoshani, O., Hirai, M., Wanga, Y., Maldonado, M., Cho, T., Boubaker, J., Swing, D. A., Tessarollo, L., Evans, S. M., Fuchs, E., & Cleveland, D. W. (2015). Chronic centrosome amplification without tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(46), E6321–E6330. https://doi.org/10.1073/pnas.1519388112

Vyas, S., Zaganjor, E., & Haigis, M. C. (2016). Mitochondria and cancer. *Cell*, *166*(3), 555–566. https://doi.org/https://doi.org/10.1016/j.cell.2016.07.002

Wong, Y. L., Anzola, J. V., Davis, R. L., Yoon, M., Motamedi, A., Kroll, A., Seo, C. P., Hsia, J. E., Kim, S. K., Mitchell, J. W., Mitchell, B. J., Desai, A., Gahman, T. C., Shiau, A. K., & Oegema, K. (2015). Reversible centriole depletion with an inhibitor of Polo-like kinase 4. *Science*, *348*(6239), 1155–1160. https://doi.org/10.1126/science.aaa5111

Chapter 5

Discussion

In this thesis, I sought to provide a comprehensive quantitative description of centriole number dynamics in proliferating cell populations by co-opting key concepts of evolutionary theory. Multiple studies have described that centriole numbers return to equilibrium after transient increases or decreases (Baudoin et al., 2020; Galofré et al., 2020; Lambrus et al., 2015; Sala et al., 2020; Wong et al., 2015). However, this was poorly understood from a quantitative perspective. Moreover, whereas the naive expectation would be that a new equilibrium would be reached under chronic centriole number perturbations, the literature was obscure in this regard. Thus, I sought to answer two fundamental questions: (1) how can the centriole number equilibrium be generated and (2) what are the consequences of chronically perturbing that equilibrium.

In chapter II, we established a parallel between centriole overproduction and selective elimination of cells carrying extra centrioles with the well-defined principle of mutation-selection balance, wherein the appearance of deleterious mutations is countered by selection. We demonstrated that simple mutation-selection models were a poor fit to a vast collection of empirical distributions of centriole numbers per cell. Instead, the data tended to display heavy-tailed distributions which could be generated by multi-step centriole number increases. Interestingly, the best-fitting models to most data sets assumed that selection was equally strong regardless of the number of extra centrioles. Overall, we inferred a high fitness penalty associated with extra centrioles and variable overproduction rates for the analysed cell lines and showed that highly accurate parameter estimates could be obtained with a modest, and experimentally feasible, increase in the number of sampled cells. In summary, we showed that mutation-selection/overproduction-selection balance was a suitable approach to analyse centriole numbers per cell at equilibrium, upon

the inclusion of biologically plausible modes of centriole biogenesis.

In chapter III, we developed a broader model to study random centriole segregation and centriole number changes along the cell cycle, in addition to centriole overproduction and selection. First, we showed that asymmetric divisions, whereby one daughter cell inherits wild-type centriole numbers and the other inherits all preexisting extra centrioles, can improve population fitness if selection against centrosome amplification acts specifically in G1, but less so if it acts in other stages. Second, we found that these divisions were more frequent for intermediate overproduction rates. Then, we studied how the timing of selection and the rate of centriole overproduction affected the equilibrium cell cycle profile. Generally, if selection against extra centrioles acted specifically in one cell cycle stage, cells tended to accumulate in that stage. As a consequence, the distribution of centriole numbers in interphase cells varied unpredictably compared to mitotic cells depending on the timing of selection. Taken together, our model indicated that random centriole segregation had little impact on fitness or relative abundance of cells with centrosome amplification and that knowledge of when selection acts along the cell cycle is important for correct assessing of centriole number anomalies in interphase.

In chapter IV, our goal was to characterise how cultured human cell populations changed over time under chronic induction of Plk4 overexpression, at different levels. We used a non-transformed human breast cell line carrying a doxycycline-inducible Plk4 transgene, which we chronically induced at different levels for over two months. We observed that the relative frequency of mitotic cells with extra centrioles decreased exponentially over time, unlike control populations lacking the transgene or carrying a truncated form of Plk4, where centriole number remained relatively constant. The rate of decrease was proportional to

the dose of doxycycline the populations experienced over the course of the experiment. We observed an initial fitness deficit in these populations, which improved over time. Indeed, fitness showed a strong anti-correlation with the relative frequency of mitotic cells with extra centrioles. Making use of proliferation dyes and DNA staining, we determined that this fitness penalty was due to slower proliferation and cell cycle delay/arrest. Plk4 was overexpressed over time in induced MCF10A-Plk4 populations, despite a temporary reduction in mRNA levels. In contrast, preliminary data suggested that Plk4 levels were initially high at the centrosome upon doxycycline treatment but eventually returned to basal levels. We concluded that chronic centriole overproduction was suppressed by strong negative selection and down-regulation of Plk4-mediated centriole overproduction. Both these processes were less efficient for cells experiencing moderate centriole overproduction.

Overall, this thesis provides crucial insights on the interaction between centriole overproduction and selection, as well as other processes, such as centriole segregation, and the resulting centriole number dynamics across generations. Our main findings can be summarised as follows: (1) equilibrium centriole number distributions can be quantitatively explained by multi-step overproduction events and uniformly strong selection; (2) as a corollary, this is supported by the observation that the relative frequency of cells with extra centrioles is a better predictor of population fitness rather than mean number of centrioles per cell; (3) centriole segregation may improve fitness when selection against extra centrioles occurs in G1; (4) asymmetric divisions under random centriole segregation can be maximised at intermediate centriole overproduction rates; (5) chronic centriole number increases following Plk4 overexpression are countered by negative selection and negative regulation of centriole overproduction.

On how centriole numbers are distributed in populations of proliferating cells

It is widely known that centriole number can vary substantially in populations of proliferating cells. However, this phenomenon was previously understudied. In chapter II, we provided the first quantitative characterisation of centriole numbers per cell in a plethora of human cell lines. We showed that variation in centriole numbers can be accurately quantified using mathematical models based on two fundamental factors: the rate of centriole overproduction/overduplication and the strength of selection against extra centrioles. In chapter III, we build upon our initial approach to explore the role of centriole segregation and centriole number changes along the cell cycle in shaping the distribution of centriole numbers per cell. Taken together, we contributed with a novel computational approach that allows centriole number variation to be quantified in the context of its underlying cellular processes.

Our approach allowed us to infer from centriole number data: (1) the probability of a cell producing a certain number of extra centrioles, (2) the fitness cost associated extra centrioles, (3) if the number of extra centrioles affects the strength of selection, (4) how extra centrioles are segregated, (5) and when selection acts against extra centrioles. The advantage of this approach is that it allows one to extract more information from standard protocols than using conventional statistical methods that simply test for differences in the level of centrosome amplification, for instance. A limitation of the models is that the predictions are not necessarily unique. For example, an increase in percentage of cells with extra centrioles might be explained by higher overproduction and/or weaker selection. As we demonstrated in chapter II, some of these limitations can be circumvented with

larger sample sizes and, as discussed in chapter III, by acquiring additional data on cell cycle profiles or interphase centriole number distributions. Although these procedures are standard practice, the usefulness of these models lies in the fact that they can be used to formulate precise quantitative hypotheses that can be experimentally tested.

How could these hypotheses be tested? For example, the number of newly assembled centrioles can be assessed using daughter centriole-specific markers, such as SAS-6 (in human cell lines at least, Strnad et al., 2007), in synchronised populations. Strength of selection can be estimated using competitive fitness assays such as the one we employed in chapter IV. The effect of centriole numbers on fitness is perhaps more challenging to quantify. In an ideal scenario, one could perform live-cell imaging and track centriole numbers and relate them of the probability of undergoing cell death, or to the duration of the cell cycle. However, live-tracking of centrioles can be challenging and very large sample sizes would be required, since one would have to gather sufficient data on cells starting with two, three, four, five... centrioles. An indirect approach relating centriole numbers in anaphase figures to spindle (multi-)polarity or relating centriole numbers in interphase cells with accumulation of nuclear p53 might be more feasible. The disadvantage of this approach neither multipolar anaphases nor p53 activation are direct measures of fitness. Segregation of extra centrioles can be assessed by live-cell imaging, as performed by Sala *et al.* (Sala et al., 2020). Finally, the timing of selection against extra centrioles can be narrowed down using a combination of live-cell imaging, cell cycle profiling using flow cytometry, viability assays, and growth curves, among others. Indeed, these methods have allowed for the identification of several pathways that suppress the proliferation of cells with extra centrioles, such as p53-dependent cell cycle arrest, interphase delay, programmed cell death following multipolar or aneuploid

cell divisions, and mitotic delay in the presence of an active spindle-assembly checkpoint. In conclusion, the experimental techniques required for testing our model predictions have been widely employed. Crucially, our models will allow future studies to establish quantitative relationships between centriole numbers and their associated cellular parameters, such as the rate of centriole overproduction, and the strength of selection on cells bearing centrosome amplification.

How cells "count" extra centrioles

In chapter II, model selection indicated that selection acted strongly against extra centrioles but, for the majority of cell lines, it was invariant with respect to the number of extra centrioles. In other words, the prediction is that mitotic cells with five, six, seven... centrioles have an equal chance of arresting proliferation or dying. Coherently, in chapter IV, we observed a linear fitness decrease with the percentage of mitotic cells carrying centrosome amplification. By comparison, mean centriole number was a worse predictor of fitness. This result provided indirect evidence of a constant fitness cost of extra centrioles.

One could expect that higher numbers of centrioles could impart a higher burden upon cells. First, the resources necessary for assembling subcellular structures should scale up with their abundance. To build a higher number of structures, a greater quantity of precursors is necessary and the energy required to precisely order them should also be higher (Lynch et al., 2022). Centrioles are relatively small structures compared to organelles such as mitochondria and plastids, or even in comparison to the total membrane volume present within the cell (Lynch et al., 2022). They might also experience a lower turnover of their components in comparison to many other subcellular features, namely the microtubule and actin cytoskeletons (Werner et al.,

2017). Thus, the cost of producing and maintaining extra centrioles might be negligible. In contrast, it might be harder to reason how certain selective mechanisms are equally effective regardless of the number of extra centrioles. It seems plausible that higher numbers of extra centrioles might more efficiently trigger p53 activation or increase the probability of multipolar divisions. In fact, this seems to be the case for cells with fewer centrioles than expected. More specifically, there is data suggesting that cytokinesis failure is likelier in mitotic cells with zero centrioles than in mitotic cells with one to three centrioles (Wong et al., 2015). More data is ultimately needed to assert if the number of extra centrioles affects cell viability.

In theory, if extra centrioles are thoroughly eliminated from proliferating cell populations, regardless of their number, it would necessitate a mechanism that recognises the presence of centrosome amplification and is both highly sensitive and highly specific. In other words, the cell must correctly assess that it contains either the "correct" number of centrioles *or* an abnormally high number. One possible mechanism for achieving this would be the existence of a molecule that can sense the presence of extra centrioles (Hyman et al., 2014). Recently, the PIDDosome has been proposed as a putative centriole number counter, purportedly based on the number of subdistal appendages/ODF2 foci or distal appendages/ANKRD6 levels, but the exact workings of this pathway have yet to be resolved (Burigotto et al., 2021; Fava et al., 2017). To achieve high sensitivity, such a sensor should generate some form of positive feedback to generate a robust response. In many signalling pathways, such as the MAPK pathway, phosphorylation cascades can lead to this sort of positive feedback, since at each step of the cascade, any given kinase can potentially activate multiple targets (Ma and Nicolet, 2023). To achieve high specificity, mechanisms such as proofreading allow the cell to "verify" a given outcome iteratively (Hopfield,

1974; Mckeithan, 1995). Thus, if an error in the sensing mechanism occurs occurs - for example, if it is activated in cells with wild-type centriole numbers - there are several steps in which it can be corrected. It will be important to address in the future if these centriole number counting mechanisms display both high sensitivity and specificity.

In the future, it will be important to investigate if the number of extra centrioles does affect the probability of cell cycle arrest or cell death and, consequentially, how the mechanisms responsible for "counting" centrioles and eliciting such outcomes operate, and how both differ between cell types. Characterising these mechanisms should also shed light on how extra centrioles are maintained and their contribution to human disease.

Sustaining chronic centrosome amplification

In chapter IV we observed that the relative frequency of cells with extra centrioles decreased over time depending on the initial degree of centrosome amplification. Although extra centrioles were highly deleterious, the observed dynamics could not be explained by continued negative selection against cells with extra centrioles. Rather, basal centriole numbers were likely recovered due to a decrease in the rate of centriole overproduction. Alternatively, since MCF10A-Plk4 populations were polyclonal, it is possible that cells which were capable of Plk4 overexpression but not centriole overproduction rose in frequency. However, we were unable to pinpoint the cause for this purported decrease in centriole overproduction.

The conditions in which chronic centrosome amplification is maintained remain elusive. *In vivo*, chronic centrosome amplification was observed only in a small subset of tissues upon chronic Plk4 overexpression (Levine et al., 2017; Vitre et al., 2015). The brain, the lung, and the kidney seem particularly non-permissive

to centrosome amplification. The liver can sustain chronic centrosome amplification, perhaps due to the fact that hepatocytes tend to naturally develop abnormal centriole numbers. In the skin epidermis, chronic centrosome amplification following Plk4 overexpression can be maintained but it might depend on the overexpression system (Coelho et al., 2015; Kulukian et al., 2015; Serçin et al., 2016). In other tissues, such as the spleen and the pancreas, maintaining chronic centrosome amplification might be related to the degree of chronic Plk4 overexpression (Levine et al., 2017; Vitre et al., 2015). Our study provides the first evidence that chronic centrosome amplification is counteracted in a human breast cell line regardless of the initial level of Plk4 overexpression.

The mechanisms that counteract centrosome amplification also appear to be context-dependent. In the brain, Plk4 overexpression and extra centrioles trigger generalised cell death, which can lead to microcephaly (Marthiens et al., 2013). In the skin epidermis, Plk4 overexpression and centrosome amplification induce defects in tissue architecture and p53-dependent or independent cell death. Knocking out p53 in the skin epidermis might be sufficient to maintain, or delay the loss of, centrosome amplification but in one study the transgenic Plk4 promoter was silenced and wild-type-like centriole numbers were recovered (Coelho et al., 2015; Kulukian et al., 2015; Serçin et al., 2016). In other tissues, such as the lung or kidney, Plk4 levels can be chronically elevated without leading to chronic centrosome amplification. Knocking out p53 in these mice rescued the loss of centrosome amplification but it also led to an increase in Plk4 expression (Vitre et al., 2015). Depending on the cell line, chronic Plk4 overexpression may lead to chronic centrosome amplification in some cases (Vitre et al., 2015) whereas in others, p53 is activated and centriole numbers are restored to their initial levels (Holland et al., 2012; Levine et al., 2017). Taken together, these results suggest that there are

two main routes for repressing chronic centrosome amplification: strong selection against cells with extra centrioles and inhibition of centriole overproduction. Our results support the idea that both these mechanisms can act simultaneously. In the future, it will be important to tease apart how Plk4 expression is regulated and test if chronic centrosome amplification can be maintained under different conditions, such as in the absence of p53 or upon overexpression of STIL or SAS-6.

On the utility of centriole numbers as cancer biomarkers

In chapters II and III, we inferred rates of centriole overproduction and strength of selection, and addressed how the relationship between centrosome amplification and the cell cycle can affect scoring of centriole number anomalies. We showed that various parameters can be inferred from centriole number data but there are conditions in which the same empirical result can be explained by different mechanisms.

The commonality of centriole number alterations in cancer and their correlation with poor clinical outcomes deems that centriole numbers can be suitable biomarkers for prognosis (Bettencourt-Dias et al., 2011; Chan, 2011; Godinho and Pellman, 2014). Recently, several protocols have been developed for quantifying centriole numbers in tumour biopsies. While encouraging, these protocols still suffer from considerable limitations - they are laborious and prone to underscoring since biopsies do not guarantee that the whole volume of a given cell will be sectioned (Lopes et al., 2018; Wang et al., 2019). Such an issue is not expected in liquid biopsies and would be interesting to assess centriole number anomalies in this scenario.

On fundamental biological terms, it is also not clear cut what

these anomalies may entail. If cells with centrosome amplification are observed, is it relevant for cancer development if they are heavily counter-selected? If high levels of centrosome amplification are observed, do cells bear an abnormally high error rate in centriole biogenesis or are they particularly tolerant to the presence of extra centrioles? Our modelling approach may help disentangle these different scenarios. Less encouragingly, since most cells in biopsies are non-mitotic (and many even quiescent), the relationship between cell cycle progression and centrosome amplification may deem that the same empirical distribution of centriole numbers per cell may be explained by different underlying processes. Since this ambiguity might be solved by supplying the models with information on the cell cycle, co-staining biopsies with proliferation markers, such as Ki67, or cell cycle markers might be sufficient to improve the quality of parameter inference. The potential of this approach for prognosis is that one would be able to relate not only the degree of centrosome amplification with disease progression or patient survival but also its causes. For example, it could be possible to discern if a given level of centrosome amplification was possibly generated due to stronger overproduction or weaker selection. The outcome for the patient might be different depending on either scenario. Therefore, our models may assist clinicians in better monitoring cancer progression. In the future, it will be important to validate these models with patient data and test if they could be a useful tool in a medical setting.

On the role of extra centrioles in cancer development

In chapter II, we inferred strong selection against extra centrioles in a variety of cell lines. In chapter IV, we observed that extra centrioles induce a high fitness cost depending on how frequently

they occurred. In other words, modest errors in centriole biogenesis, leading to few cells with centrosome amplification/low numbers of extra centrioles, can be better tolerated by proliferating cells. Furthermore, in chapter III, we demonstrated that under random centriole segregation, the probability of completely asymmetric divisions, which can facilitate chromosome missegregation, is maximised for moderate centriole overproduction rates. Based on these results, one could hypothesise that low levels of centrosome amplification might be conducive to tumourigenesis.

In the hundred-year span between Boveri's initial proposal that centrosome amplification could impact cancer development and current times, the precise contribution of extra centrioles to tumour formation remains to be fully understood. Whereas the hypothesis that cancer arises from multipolar divisions triggered by the presence of extra centrioles has been discredited (Ganem et al., 2009), some studies indicate that centrosome amplification can lead to cancer formation by inducing less severe levels of aneuploidy (Levine et al., 2017; Shoshani et al., 2021,reviewed in Bloomfield and Cimini, 2023). Overall, the main findings in this respect can be summarised as follows: strong chronic Plk4 overexpression from an early age does not induce chronic centrosome amplification in most tissues nor promote cancer development *per se* (Coelho et al., 2015; Kulukian et al., 2015; Serçin et al., 2016; Vitre et al., 2015). Low levels of chronic Plk4 overexpression can induce chronic centrosome amplification in some tissues and promote formation of specific cancer types (Levine et al., 2017). In the absence of p53 or in intestinal cancer models, chronic overexpression of Plk4 and centrosome amplification can accelerate tumour formation (Levine et al., 2017; Serçin et al., 2016). Finally, transient Plk4 overexpression is sufficient to induce the appearance of certain tumour types. In summary, there are some lines of evidence suggesting that centrosome amplification can potentiate tumour formation.

In order for extra centrioles to initiate tumourigenesis, they must first appear in the population, through genetic or environmental alterations, and be maintained for some amount of time. This can occur even if centrosome amplification is counter-selected. The efficacy of natural selection in purging deleterious mutations depends on the population size (Hill, 2002). If the population size is low, weakly deleterious mutations can appear and rise in frequency, eventually even reaching fixation. In adult tissues, only a small subset of cells are capable of proliferation. Thus, the population size of proliferating cells is presumably small. Since our results in chapter IV indicate that the fitness cost of extra centrioles depends on the degree of centrosome amplification, it seems plausible that any alterations leading to small errors in centriole biogenesis might evade selection and rise in frequency in the population. Thus, if extra centrioles *per se* can induce tumourigenesis, we expect that they should appear in pre-malignant stages at low frequencies, as is experimentally observed in Barrett's esophagus (Lopes et al., 2018).

After establishing themselves in the population, extra centrioles should induce induce some alteration which lead to malignancy. In order to be successful, these alterations should outweigh or compensate for the consequences of extra centrioles. The likeliest candidate for such an alteration is aneuploidy, since centrosome amplification favours chromosome segregation errors (Ganem et al., 2009). Live tumours are frequently aneuploid and display chromosomal instability. Gains of chromosomes can lead to the amplification of oncogenes or loss of tumour suppressors, and can be presumably targeted by positive selection. Indeed, tumours often bear recurrent aneuploidies which could have been maintained by positive selection (Klaasen and Kops, 2022; Nicholson and Cimini, 2015). Since extra centrioles can generate chromosome missegregation events quite frequently, it might not even be necessary to maintain extra centrioles for too long, since

beneficial karyotypes can be generated, in principle, in a single generation. Coherently, transient Plk4 overexpression, leading to a temporary bout of centrosome amplification, is sufficient to induce development of lymphomas and sarcomas (Levine et al., 2017; Shoshani et al., 2021). If extra centrioles are maintained in pre-malignant cell populations, as empirically observed, then our results in chapter III and IV seem to indicate that cancers are likelier to appear if the penetrance of centrosome amplification is not too severe, since it imparts a lower fitness penalty and may further enhance chromosomal instability due to asymmetric segregation of extra centrioles (Cosenza et al., 2017).

Nevertheless, the relationship between extra centrioles and cancer evolution is probably more complex. Although remarkably widespread and despite frequent alterations in gene expression of key centriolar genes, (Chan, 2011), the causes of centrosome amplification in live tumours are still not known. Since the literature has focused predominantly on centrosome amplification following Plk4 overexpression, this scenario might be both unrealistic and confounding since the observations might be partly due to Plk4 *per se*. In the future, it will be critical to identify the source of centrosome amplification in live tumours and/or induce centrosome amplification using other means to fully characterise their role in cancer development.

Another aspect to consider is the heterogeneity in centriole numbers within cancer cell populations (Bettencourt-Dias et al., 2011). It is implied in chapters II and III that this heterogeneity stems from the randomness of the biological processes leading to centriole overproduction. The data seem to be coherent with this hypothesis since our models accurately explained the observed variation in centriole numbers per cell. Conversely, this heterogeneity could explain the dynamics we observed in chapter IV, as previously mentioned. In live tumours, which tend to

accumulate significant genetic variation, it is possible that several sub-populations with different propensities for centrosome amplification exist. Our results collectively indicate that such sub-populations could be subject to different selective pressures against centrosome amplification, as inferred in chapter II and observed in chapter IV, or experience different levels of aneuploidy, as predicted in chapter III. Distinguishing between different sources of heterogeneity in centriole numbers is imperative for dissecting the role of extra centrioles in cancer development.

As previously mentioned, the benefits of centrosome amplification for a growing pre-cancerous lesion must supersede their significant fitness cost. In some tissues, such as the liver, centrosome amplification occurs naturally so this hurdle might be less impactful (Donne et al., 2021). On the other hand, for the same reason, these tissues might be better equipped to prevent malignant transformation caused by extra centrioles. In many cases, it is possible that extra centrioles are mere passengers in the oncogenic process, brought about by other abnormalities such as compromised p53 activity (observed in 53% of all cancers) or whole-genome duplication (observed in 42% of all cancers). Extra centrioles are also known to affect cell migration, increase invasiveness, and alter signalling in autocrine and paracrine fashion, so their presence might actually be beneficial at later stages of cancer formation (Arnandis et al., 2018; Godinho et al., 2014). Future work should focus on disentangling these complex interactions between centrioles and cell physiology in order to fully understand their role in cancer development.

Evolution and cell biology

In chapters II and III, we adopted a population-level approach to study the causes and consequences of centriole number variation. In chapter IV, we estimated fitness of cell populations with

and without centrosome amplification and quantified proxies for the centriole overproduction rate. This multi-faceted approach allowed us to generate novel insights on the cross-generational dynamics of cells carrying extra centrioles. To the best of our knowledge, we delivered the first comprehensive study on the relationship between the level centrosome amplification and population fitness.

Unifying evolutionary and cell biology is ultimately necessary for garnering a holistic understanding of how certain structures arise and change over time (Islas-Morales et al., 2021; Lynch et al., 2022). Whereas evolutionary theory might be more concerned on the selective pressures associated with a given genotype and/or phenotype, it tends to simplify how the genotype gives rise to a certain phenotype. Cell biology, on the other hand, is deeply invested into dissecting the complex relationship between phenotypes and genotypes, paying lesser attention to how this link might affect living organisms over time. In recent years, both disciplines have intermingled and allowed for novel research to be carried out. For cancer research, since it is, in essence, a process of somatic cell evolution combining evolutionary and cell biology approaches has allowed researchers to approach disease progression in an integrative way (Ní Leathlobhair and Lenski, 2022). On a more fundamental aspect, combining both fields has allowed researchers to ask how gene regulatory networks are tailored by mutational processes, based on the biophysics of transcription factor binding, what are the essential components for a minimal organisms and how these lifeforms evolve, and the bioenergetic constraints to gaining or losing subcellular structures (Lynch et al., 2022). In this thesis, I demonstrated how abundance of subcellular structures can seamlessly link to basic principles of theoretical populations genetics. Interdisciplinary efforts such as the one we carried out are vital for finding a common ground between cell biology and evolution.

References

Arnandis, T., Monteiro, P., Adams, S. D., Bridgeman, V. L., Rajeeve, V., Gadaleta, E., Marzec, J., Chelala, C., Malanchi, I., Cutillas, P. R., & Godinho, S. A. (2018). Oxidative Stress in Cells with Extra Centrosomes Drives Non-Cell-Autonomous Invasion. *Developmental Cell*, *47*(4), 409–424.e9. https://doi.org/10.1016/j.devcel.2018.10.026

Baudoin, N. C., Nicholson, J. M., Soto, K., Martin, O., Chen, J., & Cimini, D. (2020). Asymmetric clustering of centrosomes defines the early evolution of tetraploid cells. *eLife*, *9*. https://doi.org/10.7554/eLife.54565

Bettencourt-Dias, M., Hildebrandt, F., Pellman, D., Woods, G., & Godinho, S. A. (2011). Centrosomes and cilia in human disease. *Trends in Genetics*, *27*(8), 307–315. https://doi.org/10.1016/j.tig.2011.05.004

Bloomfield, M., & Cimini, D. (2023). The fate of extra centrosomes in newly formed tetraploid cells: Should i stay, or should i go? *Frontiers in Cell and Developmental Biology*, *11*. https://doi.org/10.3389/fcell.2023.1210983

Burigotto, M., Mattivi, A., Migliorati, D., Magnani, G., Valentini, C., Roccuzzo, M., Offterdinger, M., Pizzato, M., Schmidt, A., Villunger, A., Maffini, S., & Fava, L. L. (2021). Centriolar distal appendages activate the centrosome-piddosome-p53 signalling axis via ankrd26. *The EMBO Journal*, *40*(4), e104844. https://doi.org/https://doi.org/10.15252/embj.2020104844

Chan, J. Y. (2011). A clinical overview of centrosome amplification in human cancers. *Int. J. Biol. Sci.*, *7*(8), 1122–1144.

Coelho, P. A., Bury, L., Shahbazi, M. N., Liakath-Ali, K., Tate, P. H., Wormald, S., Hindley, C. J., Huch, M., Archer, J., Skarnes, W. C., Zernicka-Goetz, M., & Glover, D. M. (2015). Over-expression of Plk4 induces centrosome amplification, loss of primary cilia and associated tissue hyperplasia in the

mouse. *Open Biology*, *5*(12). https://doi.org/10.1098/rsob.150209

Cosenza, M. R., Cazzola, A., Rossberg, A., Schieber, N. L., Konotop, G., Bausch, E., Slynko, A., Holland-Letz, T., Raab, M. S., Dubash, T., Glimm, H., Poppelreuther, S., Herold-Mende, C., Schwab, Y., & Krämer, A. (2017). Asymmetric centriole numbers at spindle poles cause chromosome missegregation in cancer. *Cell Reports*, *20*(8), 1906–1920. https://doi.org/https://doi.org/10.1016/j.celrep.2017.08.005

Donne, R., Sangouard, F., Celton-Morizur, S., & Desdouets, C. (2021). Hepatocyte polyploidy: Driver or gatekeeper of chronic liver diseases. *Cancers*, *13*(20). https://doi.org/10.3390/cancers13205151

Fava, L. L., Schuler, F., Sladky, V., Haschka, M. D., Soratroi, C., Eiterer, L., Demetz, E., Weiss, G., Geley, S., Nigg, E. A., & Villunger, A. (2017). The PIDDosome activates p53 in response to supernumerary centrosomes. *Genes and Development*, *31*(1), 34–45. https://doi.org/10.1101/gad.289728.116

Galofré, C., Asensio, E., Ubach, M., Torres, I. M., Quintanilla, I., Castells, A., & Camps, J. (2020). Centrosome reduction in newly-generated tetraploid cancer cells obtained by separase depletion. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-65975-1

Ganem, N. J., Godinho, S. A., & Pellman, D. (2009). A mechanism linking extra centrosomes to chromosomal instability. *Nature*, *460*(7252), 278–282. https://doi.org/10.1038/nature08136

Godinho, S. A., & Pellman, D. (2014). Causes and consequences of centrosome abnormalities in cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1650). https://doi.org/10.1098/rstb.2013.0467

Godinho, S. A., Picone, R., Burute, M., Dagher, R., Su, Y., Leung, C. T., Polyak, K., Brugge, J. S., Théry, M., & Pellman, D. (2014). Oncogene-like induction of cellular invasion from centrosome amplification. *Nature*, *510*(7503), 167–171. https://doi.org/10.1038/nature13277

Hill, W. (2002). *The Mathematical Theory of Selection, Recombination and Mutation. R. Burger* (Vol. 79). Wiley.

Holland, A. J., Fachinetti, D., Zhu, Q., Bauer, M., Verma, I. M., Nigg, E. A., & Cleveland, D. W. (2012). The autoregulated instability of polo-like kinase 4 limits centrosome duplication to once per cell cycle. *Genes Dev.*, *26*(24), 2684–2689.

Hopfield, J. J. (1974). Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(10), 4135–9. https://doi.org/10.1073/pnas.71.10.4135

Hyman, A. A., Weber, C. A., & Jülicher, F. (2014). Liquid-liquid phase separation in biology [PMID: 25288112]. *Annual Review of Cell and Developmental Biology*, *30*(1), 39–58. https://doi.org/10.1146/annurev-cellbio-100913-013325

Islas-Morales, P. F., Jime Nez-Garcia, L. F., Mosqueira-Santillan, M., & Voolstra, C. R. (2021). Evolutionary cell biology (ECB): Lessons, challenges, and opportunities for the integrative study of cell evolution. *J. Biosci.*, *46*(1).

Klaasen, S. J., & Kops, G. J. P. L. (2022). Chromosome inequality: Causes and consequences of non-random segregation errors in mitosis and meiosis. *Cells*, *11*(22). https://doi.org/10.3390/cells11223564

Kulukian, A., Holland, A. J., Vitrec, B., Naika, S., Cleveland, D. W., & Fuchsa, E. (2015). Epidermal development, growth control, and homeostasis in the face of centrosome amplification. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(46), E6311–E6320. https://doi.org/10.1073/pnas.1518376112

Lambrus, B. G., Uetake, Y., Clutario, K. M., Daggubati, V., Snyder, M., Sluder, G., & Holland, A. J. (2015). P53 protects against genome instability following centriole duplication failure. *Journal of Cell Biology*, *210*(1), 63–77. https://doi.org/10.1083/jcb.201502089

Levine, M. S., Bakker, B., Boeckx, B., Moyett, J., Lu, J., Vitre, B., Spierings, D. C., Lansdorp, P. M., Cleveland, D. W., Lambrechts, D., Foijer, F., & Holland, A. J. (2017). Centrosome Amplification Is Sufficient to Promote Spontaneous Tumorigenesis in Mammals. *Developmental Cell*, *40*(3), 313–322.e5. https://doi.org/10.1016/j.devcel.2016.12.022

Lopes, C. A., Mesquita, M., Cunha, A. I., Cardoso, J., Carapeta, S., Laranjeira, C., Pinto, A. E., Pereira-Leal, J. B., Dias-Pereira, A., Bettencourt-Dias, M., & Chaves, P. (2018). Centrosome amplification arises before neoplasia and increases upon p53 loss in tumorigenesis. *Journal of Cell Biology*, *217*(7), 2353–2363. https://doi.org/10.1083/jcb.201711191

Lynch, M., Schavemaker, P. E., L., T. J., Y., Hao, & A., P. (2022). Evolutionary bioenergetics of cilliates. *Journal of Eukaryotic Microbiology*, *69(5):e12934*. https://doi.org/https://doi.org/10.1111/jeu.12934

Ma, Y., & Nicolet, J. (2023). Specificity models in mapk cascade signaling. *FEBS Open Bio*, *13*(7), 1177–1192. https://doi.org/https://doi.org/10.1002/2211-5463.13619

Marthiens, V., Rujano, M. A., Pennetier, C., Tessier, S., Paul-Gilloteaux, P., & Basto, R. (2013). Centrosome amplification causes microcephaly. *Nature Cell Biology*, *15*(7), 731–740. https://doi.org/10.1038/ncb2746

Mckeithan, T. W. (1995). *Kinetic proofreading in T-cell receptor signal transduction (protein-tyrosine kinase/major histocompatibility complex/mathematical model)* (tech. rep.).

Ní Leathlobhair, M., & Lenski, R. E. (2022). Population genetics of clonally transmissible cancers. *Nature Ecology & Evolution*,

6(8), 1077–1089. https://doi.org/10.1038/s41559-022-01790-3

Nicholson, J. M., & Cimini, D. (2015). Link between aneuploidy and chromosome instability. *International Review of Cell and Molecular Biology*, *315*, 299–317. https://doi.org/10.1016/bs.ircmb.2014.11.002

Sala, R., Farrell, K., & Stearns, T. (2020). Growth disadvantage associated with centrosome amplification drives population-level centriole number homeostasis. *Molecular Biology of the Cell*, mbc.E19–04–0195. https://doi.org/10.1091/mbc.e19-04-0195

Serçin, Ö., Larsimont, J. C., Karambelas, A. E., Marthiens, V., Moers, V., Boeckx, B., Le Mercier, M., Lambrechts, D., Basto, R., & Blanpain, C. (2016). Transient PLK4 overexpression accelerates tumorigenesis in p53-deficient epidermis. *Nature Cell Biology*, *18*(1), 100–110. https://doi.org/10.1038/ncb3270

Shoshani, O., Bakker, B., de Haan, L., Tijhuis, A. E., Wang, Y., Kim, D. H., Maldonado, M., Demarest, M. A., Artates, J., Zhengyu, O., Mark, A., Wardenaar, R., Sasik, R., Spierings, D. C. J., Vitre, B., Fisch, K., Foijer, F., & Cleveland, D. W. (2021). Transient genomic instability drives tumorigenesis through accelerated clonal evolution. *Genes Dev.*, *35*(15-16), 1093–1108.

Strnad, P., Leidel, S., Vinogradova, T., Euteneuer, U., Khodjakov, A., & Gönczy, P. (2007). Regulated HsSAS-6 Levels Ensure Formation of a Single Procentriole per Centriole during the Centrosome Duplication Cycle. *Developmental Cell*, *13*(2), 203–213. https://doi.org/10.1016/j.devcel.2007.07.004

Vitre, B., Holland, A. J., Kulukian, A., Shoshani, O., Hirai, M., Wanga, Y., Maldonado, M., Cho, T., Boubaker, J., Swing, D. A., Tessarollo, L., Evans, S. M., Fuchs, E., & Cleveland, D. W. (2015). Chronic centrosome amplification without

tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(46), E6321–E6330. https://doi.org/10.1073/pnas.1519388112

Wang, M., Knudsen, B. S., Nagle, R. B., Rogers, G. C., & Cress, A. E. (2019). A method of quantifying centrosomes at the single-cell level in human normal and cancer tissue. *Molecular Biology of the Cell*, *30*(7), 811–819. https://doi.org/10.1091/mbc.E18-10-0651

Werner, S., Pimenta-Marques, A., & Bettencourt-Dias, M. (2017). Maintaining centrosomes and cilia. *Journal of Cell Science*, *130*(22), 3789–3800. https://doi.org/10.1242/jcs.203505

Wong, Y. L., Anzola, J. V., Davis, R. L., Yoon, M., Motamedi, A., Kroll, A., Seo, C. P., Hsia, J. E., Kim, S. K., Mitchell, J. W., Mitchell, B. J., Desai, A., Gahman, T. C., Shiau, A. K., & Oegema, K. (2015). Reversible centriole depletion with an inhibitor of Polo-like kinase 4. *Science*, *348*(6239), 1155–1160. https://doi.org/10.1126/science.aaa5111

Acknowledgements

Taking on a PhD project has been one of the longest journeys in my life. I feel that adequately acknowledging all the people and all the events which led me to this point is impossible. Nevertheless, here is a relatively unstructured attempt at doing so, in full conscience that whichever words I write will probably not be enough to appreciate what so many people did for me.

First, I would like to acknowledge the IGC for housing me for the past nearly eight years, first as a master student and then as a PhD student. I am deeply grateful that I was selected as a part of the IBB 2017 cohort. In would also like to acknowledge the IBB direction and its many faces over the years for their support and handling of the PhD programme. I would also like to acknowledge the University of Bern for being my home away from home in my recent stay in Switzerland.

I would like to acknowledge the Cell Cycle Regulation and Evolutionary Dynamics/Theoretical Ecology and Evolution labs for receiving me and helping me grow as a scientists. I am deeply grateful to Mónica Bettencourt-Dias and Claudia Bank for accepting to co-supervise my PhD work and for giving me the freedom to develop my own project. Thank you for encouraging me to participate in international conferences, workshops, reviewing papers, and most importantly (for me) to take on summer students, in what I consider was a complete and thorough PhD experience.

To my colleagues in both labs, either in Oeiras or Bern, thank you for your companionship and for your support over the years, be it personal or in science-related matters. In the Cell Cycle Regulation lab, thank you to Ana Rita Marques, Camila Mariano, Carla Lopes, Catarina Nabais, Catarina Peneda, Daisuke Ito, Gaëlle Marteil, Inês Afonso, Inês Gomes, Irina Fonseca, Joana Bugalhão, Joana Ferreira, Ksenia Volkova, Leonor Nunes, Luca

Cirino, Mafalda Pimentel, Mariana Faria, Miguel Pereira, Nuria Moreno, Patrícia Rodrigues, Paulo Duarte, Pilar Ramos, Sascha Werner, Sihem Zitouni, Sónia Pereira, Sophie Dias, Swadhin Jana, Tânia Perestrelo, and others. A special thanks to Catarina Peneda for her enormous help and support in developing the experimental part of my project. In the Evolutionary Dynamics/Theoretical Ecology and Evolution labs, thank you to Adamandia Kapopoulou, Alexandre Blanckaert, Ana-Hermina Ghenu, Ana Morales-Arce,André Amado, Christian Diwo, Dave McLeod, Davide Cusseddu, Inês Fragata, Dragan Stajic, Emma Berdan, Li Juan, Julio Ayala, Jérôme Stäheli, Loïc Marrec, Loraine Hablützel, Mark Schmitz, Massimo Amicone, Russ Jasper, and Suman Das, and others. To my brilliant summer students, Dinis Seward, Helena Pericão, and Matilde Palmeira, it was a privilege to have met you and to have learned from you.

I would like to thank my thesis committee, Jorge Carneiro and Lars Jansen, for their constructive criticism on my project. A special thanks to Jorge for all he taught me as my master's supervisor.

To my school teachers and university professors, thank you for inspiring my curiosity and for your contribution to where I am now.

To my IGC 2017 cohort, Ana-Hermina Ghenu, Anastasiia Lozovska, André Madaleno, André Dias, Anton Kermanov, Camila Ramos, Catarina Carmo, Hugo Barreto, Margarida Araújo, Sahar Tehrani, Sebastiaan van den Berg, Tânia Paulo and Temitope Etibor, thank you for everything over the years. I am honoured to have shared this experience with all of you. To all IGCers, I am thankful for all the friendships I made along the way. A special mention to the IGC band, and a shout-out to all the noise we made! Thank you to all the IGC PhD Delegates for their continued work over the years - I am proud for having been a part of this group. Finally, a big thanks to the Building Bridges

in Biology organising committee - we pulled off an amazing symposium!

On a personal note, I would like to thank my family for a whole lifetime of support. In particular, I would like to thank my mother, 'salinda, and my sisters, Maria and Carolina, and my grandparents, "avô João" and "avó Linda". I could think for a million years and I would not be able to find the words for you. Just: thank you.

To my other family, who accompanied during my university studies: André, Bia, Camila, Camilo, Diana, "Hotspot", Inês, Joana, João (Diogo), João (Simão), Mameri, Mariana, Miguel, Pedro, Nuno, Rúben, Tiago, and others, thank you for everything! To my other friends, in Faro or in Alcochete, or even in Monte Abraão, thank you for having been part of my life, no matter the distance or time.

ITQB NOVA, Avenida da República, 2780-157 Oeiras, Portugal | Tel +351 214 469 100 | itqb.unl.pt