The role of space in homeostasis and preneoplasia in stratified squamous epithelia

Vasiliki Kostiou

Sidney Sussex College, University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

December 2019

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit of 60,000 words for the Clinical Medicine and Clinical Veterinary Medicine Degree Committee. These limits exclude figures, photographs, tables, appendices and bibliography.

The role of space in homeostasis and preneoplasia in stratified squamous epithelia

Vasiliki Kostiou

A major subject of study in biological research is the dynamics of stem cells in squamous epithelia. Given that most common human cancers develop from epithelia, understanding the rules of cell fate decision in these systems is key to explaining not only healthy tissue growth and maintenance but also the processes of mutagenesis and cancer. The aim of my project was to investigate the dynamics in squamous epithelial tissues both in homeostasis and preneoplasia, using cellular automata (CA) models.

Stem cell dynamics has been shown to be accurately described by a simple mathematical model, the single progenitor (SP) model. Reliable parameterisation of this model would give access to valuable quantitative information on epithelial tissue maintenance and enable investigating how mutations affect tissue dynamics. I initially identified the most appropriate method for accurately parameterising the homeostatic system.

I then sought to account for the spatial patterning of cells by implementing the SP model in two-dimensional space. The spatial model was able to reproduce the key signatures of homeostatic dynamics, thus showing that restrictions imposed by tissue organization do not alter the neutral dynamics.

Furthermore, I studied non-homeostatic dynamics in stratified squamous epithelial tissues by spatially modelling the growth and competition of non-neutral mutations as well as the effects of wounding in the tissue. The studied dynamics of Notch and p53 mutant clones in mouse epithelia has been found to be highly distinct, with the former fully colonizing the tissue whereas the latter only partially. I demonstrated that the two mutants' tissue takeover dynamics can be recapitulated by two distinct spatial feedback rules, on the basis of response to crowding, providing a mechanistic explanation of the observed distinct growth modes.

Finally, mutant competition was explored. A striking effect resulting from the spatial interaction of the two mutations in a wild-type background is that the p53 mutant cell population was always outcompeted by the Notch mutant population and appeared to shrink. Considering this consistent emergent behaviour in the competition simulations and given the paucity of Notch mutations in human cancer datasets, it is tempting to speculate that the aggressive fitness of Notch may offer a tumour-protective effect.

I would like to dedicate this thesis to my family.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Ben Hall for offering me the opportunity to pursue my PhD in his lab. Without his close mentoring and persistent support over the last four years, this thesis would not have been possible. His motivation and enthusiasm about research have deeply inspired me.

A big thank you to all Hall group members, Dr. David Shorthouse, Laure Talarmain, Michael Hall and Cassandra Kosmidou, for all the great moments we shared and for creating such a fun and friendly working environment. I really appreciate your support, your input and our interesting discussions. It was a great pleasure working with you. David, Michael, thank you for proof-reading this thesis. Special thanks go to David, for his continuous assistance, for the inspirational discussions about science and his insightful comments. I am truly grateful to Cassandra and Laure, for listening to my concerns and keeping me motivated during writing.

I would also like to extend my thanks to Dr. Phil Jones and his lab for their collaboration and for providing their data.

Last but not least, I am very much thankful to my parents Dimitris and Maria, to Ioanna and Ilias, for their endless support during these years and for encouraging me to believe in myself.

Table of contents

List of figures

List of tables

List of Abbreviations and Acronyms

2D   Two Dimensional

2xSC Two Stem Cell

ABC   Approximate Bayesian Computation

BCC   Basal Cell Carcinoma

CA   Cellular Automata

CBM Centre Based Models

CFP   Cyan Fluorescent Protein

CI   Confidence Interval

CP   Cellular Potts

Cre   Cyclization recombinase

cSCC cutaneous Squamous Cell Carcinoma

CTMC continuous time Markov chain

DAPI 4',6-diamidino-2-phenylindole

DNA   Deoxyribonucleic Acid

DNM Dominant Negative form of Maml1

ECDF Empirical Cumulative Distribution Function

EdU   5-ethynyl-2'-deoxyuridine

ER   Estrogen Receptor

GFP   Green Fluorescent Protein

GORD   Gastrooesophageal reflux disease

H2B-GFP Histone H2B-green fluorescent protein

IFE   Interfollicular Epidermis

IQR   Interquartile range

KDE   Kernel Density Estimate

KS test Kolmogorov-Smirnov test

LGCA   Lattice Gas Cellular Automata

LoxP   locus of X-over of P1

Maml1 Mastermind-like 1

MLE   Maximum Likelihood Estimation

NMSC   Non-melanoma skin cancer

OAC   Oesophageal Adenocarcinoma

ODE   Ordinary Differential Equation

OE   Oesophageal Epithelium

OSCC   Oesophageal Squamous Cell Carcinoma

RFP   Red Fluorescent Protein

SC-CP Stem Cell - Committed Progenitor

SD   Standard deviation

SMC-ABC Sequential Monte Carlo - Approximate Bayesian Computation

SP   Single Progenitor

SSA   Stochastic Simulation Algorithm

TP53   Tumour protein p53

UV   ultraviolet

WT   Wild Type

YFP   Yellow Fluorescent Protein

Chapter 1

Introduction

Summary

In this chapter, I provide the necessary background information related to my PhD project. My main research question was to understand homeostatic and preneoplastic dynamics in stratified squamous epithelial tissues. To this end, I sought to explore how epithelial cells interact and compete and how mutant cell populations expand within a tissue. I initially describe the morphological properties of stratified squamous epithelial tissues, how their dynamics can be quantified and different proposed models of epithelial maintenance. Subsequently, I discuss the effects of spatial constraints on cell dynamics imposed by tissue architecture, highlighting the importance of explicitly considering space when studying the processes of mutagenesis and tumour formation. Finally, I review the recent literature that investigates the effects of genetic mutations on epithelial homeostasis and discuss the aims of my research project.

1.1   Introduction

A primary focus of biological science is to understand complex phenomena, such as tissue growth and maintenance. These processes involve the cooperation of individual cells through a series of mechanisms which operate over different time and length scales, to generate and preserve a more complex, hierarchical structure.

Epithelial tissues protect and line the body's surfaces and cavities, serving as an interface between the body and the environment. Despite the heterogeneity among them, a common characteristic is that they consist of densely packed sheets of cells which can be organised

in a single layer or multiple layers (Donati and Watt, 2015). Notably, the majority of adult human cancers (about 90%) develop from epithelia (Frank, 2007). Two well studied epithelial tissues of particular interest are the mammalian oesophageal epithelium and the epidermis. These are stratified squamous (i.e. consisting of cells with flattened cell shape) epithelial tissues that exhibit a rapid turnover and are maintained through continuous stem cell division and differentiation. Stem cells communicate and interact with one another and their environment in a tightly coordinated manner to respond to tissue needs and ensure homeostasis. Deregulation of these fundamental processes may result in serious pathological conditions, such as cancer. Therefore, maintenance of these tissues is critically important to understand both health and cancer progression and remains a major subject of study in biomedicine.

Epithelial tissue maintenance has been studied over the last decades and different models describing homeostatic dynamics have been proposed. An increasing body of work based on lineage tracing demonstrated that stratified squamous epithelia follow a stochastic, population asymmetric self-renewal. That is to say, cell proliferation and loss remain balanced on a population basis (Clayton et al., 2007, Doupé et al., 2010, 2012, Mascré et al., 2012, Lim et al., 2013, Rompolas et al., 2016, Sada et al., 2016, Sanchez-Danes et al., 2016). However, whether there is a single or multiple underlying populations of dividing cells remains a matter of debate. Currently, there are three proposed models of stratified squamous epithelial stem cell maintenance. Among them, the simplest model is the single progenitor (SP), which supports the existence of a single dividing cell population (Clayton et al., 2007). A second model is the stem cell - committed progenitor (SC-CP), proposing a hierarchy of a slow-cycling stem cell population that gives rise to fast-cycling committed progenitors (Mascré et al., 2012). An alternative hypothesis is the two independent stem-cell (2xSC) model, which argues in favour of two stem cell populations, with different division rates (Sada et al., 2016).

The above-mentioned stochastic models describe the evolution of cell populations. However, cells grow and compete within a tissue that has a certain organisation and structure. As a consequence, the tissue imposes spatial constraints on cell movement and growth, thus affecting growth dynamics. There are several lines of evidence highlighting that spatial constraints have a strong impact on how mutant cells expand and how tumours evolve (van der Heijden et al., 2019, Noble et al., 2019, Chkhaidze et al., 2019, Lynch et al., 2017, Hall et al., 2019). Therefore, the study of space contribution to clonal interaction and growth within tissues is key to understand the progression of mutant clones towards cancer.

1.2   Stratified squamous epithelia

The role of epithelial tissues is to protect and line the body's surfaces and cavities, serving as an interface between the body and the environment. Stratified squamous epithelia are tissues consisted of densely packed sheets of squamous cells, organised in multiple layers, with the bottom layer sitting on a basement membrane. Of them, two very well studied systems are the oesophageal epithelium and the interfollicular epidermis (IFE).

The oesophagus is a tube connecting the stomach with the external environment and is lined with a vertically stratified squamous epithelium. The oesophageal epithelium is a rapidly regenerative tissue which consists of layers of keratinocytes arranged on top of a basement membrane (Figure 1.1 bottom) (Seery, 2002, Doupé et al., 2012, Alcolea, 2017). Despite sharing a common cellular hierarchy, there are fundamental histological differences between humans and mice. In humans, the oesophageal epithelium is folded along papillae which divide the epithelium into papillary and interpapillary zones, contains more layers and it is non-keratinized (Seery, 2002). Murine oesophageal epithelium holds a less complex organization. In contrast to other epithelia, it has a uniform architecture, lacking any glands or other appendages and is keratinized. Keratinization (or cornification) is the process of terminal differentiation during which keratinocytes lose their organelles and are transformed into a tough layer of flattened, dead cells (Candi et al., 2005). Keratinization of the murine oesophageal epithelium may form an extra protective barrier against abrasive food intake (Rosekrans et al., 2015). The remarkable architectural simplicity along with the rapid turnover makes murine oesophageal epithelium an ideal experimental system for studying cell dynamics.

A tissue with a constant turnover and a slightly more complex architecture is the interfollicular epidermis. Similarly to the oesophageal epithelium, the epidermis is a well characterized system and therefore considered a good prototype tissue for studying cellular behaviour. Mammalian epidermis serves as an essential barrier having a protective function on animal survival from the external environmental stresses such as dehydration, wounds and microbial infection. It is maintained by a stratified squamous epithelium which consists of layers of keratinocytes interspersed with hair follicles and sweat glands (Figure 1.1 top)(Lu et al., 2012, Page et al., 2013, Blanpain and Fuchs, 2009). Progenitor cells are located at the deepest basal layer and differentiating ones are progressively moving through the supra basal layers towards the surface where they leave the tissue (Clayton et al., 2007, Doupé et al., 2010).
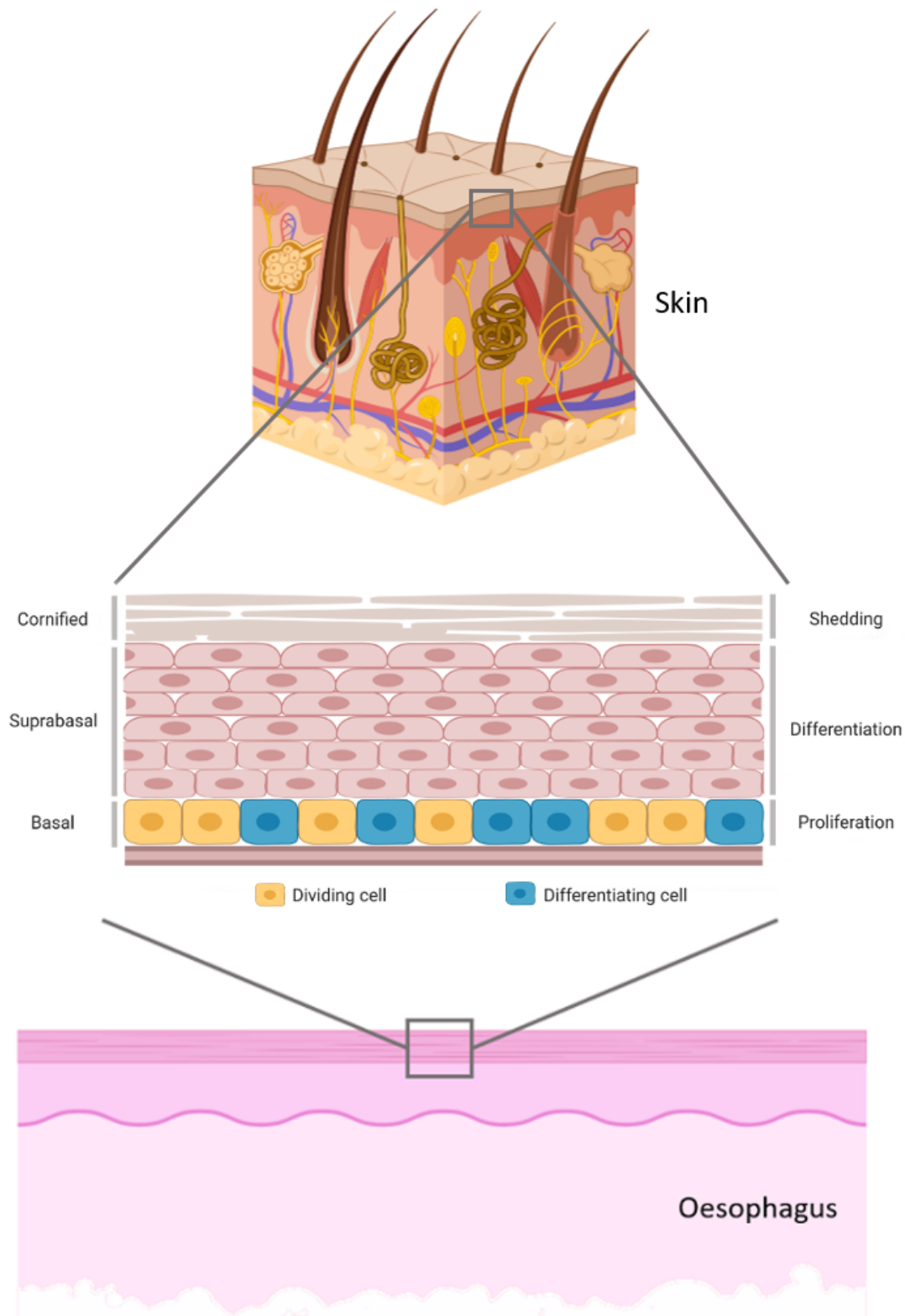
Fig. 1.1 The architecture of stratified squamous epithelial tissues of interfollicular epidermis (skin) and oesophagus in adult mice. Proliferation is restricted to the basal layer. Upon differentiation, basal cells exit the cell cycle and migrate through suprabasal layers, until eventually, they are shed from the tissue. Cell production and loss should be perfectly balanced so that homeostasis and tissue proper function is achieved. Figure based on Piedrafita et al. (2020).

Both oesophageal epithelium and epidermis share a common maintenance mechanism. Proliferation is confined to the most basal layer, where cells are arranged with their nuclei perpendicular to the membrane (Rosekrans et al., 2015). On commitment to differentiation, basal cells leave the lowermost basal layer and stratify to the upper suprabasal layers. During the differentiation process they undergo changes in their shape and orientation. In the late stages of differentiation they lose their nuclei, become large, flattened and organised parallel to the basement membrane. Eventually, they reach the tissue surface from which they are shed (Figure 1.1 middle). In a homeostatic state, the proliferation of cells in the basal layer is strongly linked to the shedding of the terminally differentiated cells (Clayton et al., 2007, Doupé et al., 2010, 2012). The mechanisms that control the balance between the production of proliferating and differentiating cells are highly important, as an excess of either proliferating or differentiating cells may result in tumourigenesis and tissue failure respectively (Doupé and Jones, 2013, Alcolea et al., 2014).

1.3   Quantifying tissue dynamics

Recent advances in experimental techniques have shed more light on our understanding of stem cell behaviour in epithelial tissues. Genetic lineage tracing provides the opportunity to label a cell type of interest and observe the fate of its progeny at different time points. The major advantage of this approach is that the behaviour of cells is studied in their native environment in contrast to other methods such as tissue culture or transplantation assays (Alcolea and Jones, 2014). Cells within tissues are marked with an inheritable fluorescent label which is passed to their descendants each time a division occurs. Thus, fate decisions can be recorded over time.

To achieve cell labelling, mice are genetically modified to express a bacterial recombinase enzyme Cre in the cell type of interest. The Cre enzyme is usually fused with a mutant estrogen receptor (ERT). Upon treatment with a drug (i.e tamoxifen), creERT fusion protein enters the nucleus and enables recombination in its recognition site, loxP. More specifically, Cre activity leads to the expression of a reporter gene, usually a fluorescent protein such as GFP, by excising a "stop" cassette that blocks transcription of the reporter (Figure 1.2a). If the drug is induced at low doses, targeted cells are labelled at low frequency, resulting in cell families deriving from a single ancestor (cell clones) (Figure 1.2b)(Alcolea and Jones, 2014). To further avoid clone fusion, multicolour reporters (e.g. confetti systems (Snippert et al., 2010), RGB marking with lentiviral vectors (Weber et al., 2012)) have been also used, where cells are labelled randomly with more than one colour.

Traditional lineage tracing allows the visualisation of static "snapshots" of clonal behaviour at certain timepoints whereas more modern live imaging techniques enable the live tracking of fate choices of individual cells within labelled clones (Park et al., 2016). The distribution of cell clone sizes observed in lineage tracing experiments can be used to further investigate clonal dynamics across the population of cells and make important observations regarding cellular behaviour within tissues.

a

b

Time Chase
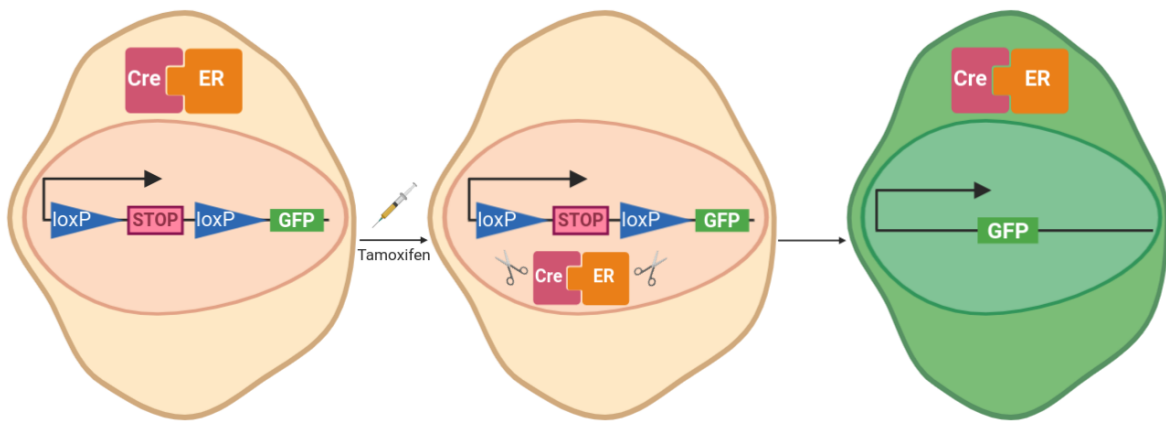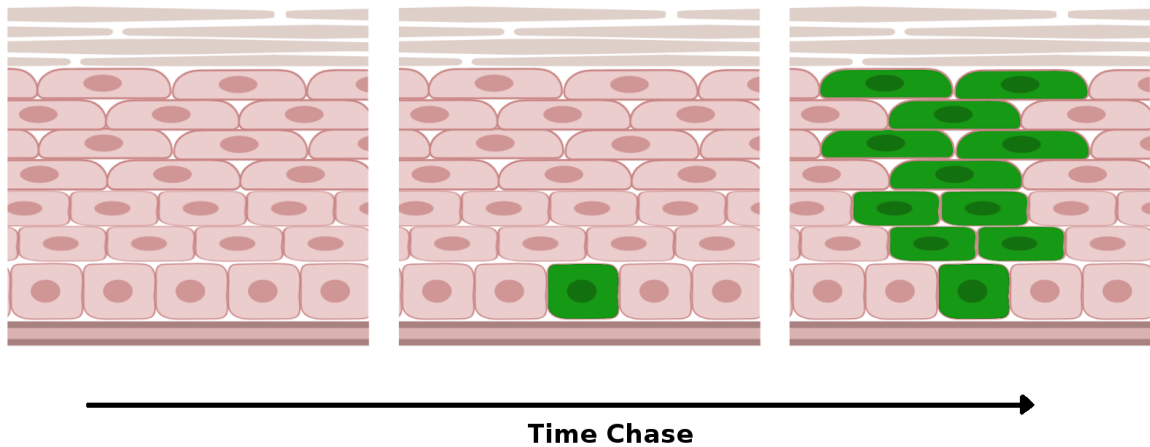
Fig. 1.2 Transgenic lineage tracing. a) Genetically modified mice express a bacterial recombinase enzyme Cre which is fused with a mutant estrogen receptor (ERT) and is only active following treatment with a drug (Tamoxifen). After treatment, CreER fusion protein enters the nucleus, removes a loxP flanked "stop" cassette and leads to the expression of a reporter gene, here a GFP fluorescent protein. The progeny of the labelled cell can be tracked as they will also express the fluorescent reporter. (b) Schematic representation showing a time chase of a typical labelled clone. If targeted cells are labelled at low frequency, single-cell derived clones may arise. Figure adapted from Alcolea and Jones (2013).

1.4 Models of epithelial homeostasis

Maintenance in stratified squamous epithelia has been studied for several decades and several attempts have been made to describe the way cells behave to maintain and renew these stratified squamous tissues, resulting in controversial conclusions. An early attempt was made by Leblond and colleagues in rat oesophageal epithelium where using pulse labelling with tritiated thymidine allowed the monitoring of cell fates. Leblond and colleagues observed that 48 hours post labelling, 58% of stem cell divisions gave rise to two daughters that either remained at the basal layer or moved to the suprabasal layer and 42% produced one daughter remaining at the basal layer and one moving at the suprabasal layer (Marques-Pereira and Leblond, 1965). These findings suggested that homeostatic dynamics in the oesophageal epithelium is maintained by a single population of stem cells.

Subsequent studies however, both in mouse oesophagus and epidermis, proposed an alternative model of tissue homeostasis where long-lived, slow cycling basal stem cells divide asymmetrically to give rise to short-lived rapidly cycling transit amplifying (TA) cells that execute the appropriate differentiation pathways after several rounds of division (Mackenzie, 1970, Potten and Booth, 2002, Croagh et al., 2007, Kalabis et al., 2008). The identification of distinct oesophageal stem cell subpopulations was mainly based on the expression of different levels of cell surface markers. Proliferative heterogeneity was believed to be further evidenced by tissue culture experiments of human epidermal keratinocytes where some cells were able to form rapidly expanding and persistent colonies (holoclones), potentially representing stem cells, whilst others generated small colonies with limited proliferative potential that became lost through terminal differentiation (paraclones), potentially representing TA cells (Barrandon and Green, 1987). The stem/TA hypothesis argues in favour of an invariant asymmetric mode of self renewal, where long term dynamics is expected to result in similar sized clonal units.

A series of recent studies based on genetic lineage tracing further questioned the invariant asymmetry paradigm and proposed a model of oesophageal and epidermal homeostasis in mouse that supports the early Leblond's findings (Clayton et al., 2007, Doupé et al., 2010, 2012, Lim et al., 2013, Rompolas et al., 2016). If a stem/TA mode of maintenance was true, stem cell clones would have been expected to initially expand and eventually adopt an invariant size whereas clones derived from TA cells would have been expected to become extinct following an initial limited growth. On the contrary, quantitative analysis of clonal cell fate behaviour in those studies revealed that the number of clones decreased over time but the size of the remaining ones progressively increased to ensure an overall constant

labelled cell population. These observations argue in favour of a population asymmetric self-renewal, that is, the outcome of each individual division is stochastic but an overall balance between cell production and loss is maintained. Thus, the proposed model supported the existence of functional equivalence within stem cell population, suggesting that a single progenitor population with balanced stochastic fate maintains epithelial homeostasis (the single progenitor (SP) model).

Further studies, also based on genetic lineage tracing, suggested that the observed population asymmetrical self-renewal strategy is underpinned by proliferative heterogeneity and proposed the hierarchical stem cell - committed progenitor (SC-CP) model. Mascré et al. (2012), Sanchez-Danes et al. (2016) used two targeted promoters (Involucrin and Keratin 14) to induce genetic labelling in mouse tail epidermal cells. By performing quantitative clonal analysis, they claimed that clones derived from the two different genetic constructs had distinct dynamics, with the keratin 14-creER targeted cells exhibiting an increased survival rate, and thus were thought to represent a stem cell subpopulation. Based on these findings, Mascré et al. (2012) and Sanchez-Danes et al. (2016) proposed that epidermal homeostatic dynamics can be described by a proliferative hierarchy. The hierarchical model posits the existence of two distinct populations, slow-cycling stem cells and quickly-cycling progenitor cells that both undergo population asymmetric self renewal. Stem cells divide stochastically at a slower rate, to generate three possible fate outcomes with balanced probabilities: two stem cells, one stem and one progenitor cell or two progenitor cells. This process gives rise to committed progenitors that divide at a faster rate and follow a similar stochastic division pattern generating either two progenitor cells, two differentiating cells or one daughter of each type.

Aside from the SP and SC-CP models, a third population asymmetric model of epithelial homeostasis was also proposed based on lineage tracing data in mouse back skin. According to that model, there are two independent populations of stem cells that divide at two different rates (2xSC model), following a stochastic fate choice pattern (Sada et al., 2016). The 2xSC hypothesis was based on the analysis of cell division kinetics from H2BGFP dilution experiments.

1.5   The Single Progenitor (SP) model

Quantitative analysis across several studies based on genetic lineage tracing in mouse skin and oesophagus over the past years revealed that homeostasis in these tissues can be accurately recapitulated by a stochastic process with some studies suggesting a single underlying stem

cell population (Clayton et al., 2007, Doupé et al., 2010, Klein and Simons, 2011, Doupé et al., 2012, Lim et al., 2013, Rompolas et al., 2016), whilst others are proposing the existence of distinct populations (Mascré et al., 2012, Sanchez-Danes et al., 2016, Sada et al., 2016). In this thesis I shall focus on the simplest proposed model, the single progenitor (SP) model.

This model supports the idea that the tissue is maintained by a single, equipotent progenitor population of basal cells that are able to give rise to either stem cells or differentiating daughters stochastically. Each division leads to three possible fate outcomes: two proliferating, two differentiating or one proliferating and one differentiating daughter may arise. The probabilities of each type of division are balanced, allowing homeostasis to be achieved (Figure 1.3a).

The stochastic nature of this tissue renewal paradigm implies that cell clones compete neutrally within the tissue, with some of them being "lucky" by chance and expand as they undergo symmetric division, whilst others shrink and become extinct as they undergo symmetric differentiation (Figure 1.3b).

This population asymmetric mode of renewal can be further specified by a set of parameters and can be described as a continuous time Markovian process, as shown by Doupé et al. (2012), Clayton et al. (2007):

$$
A \xrightarrow{\lambda}
\begin{cases}
AA & r \\
AB & 1 - 2r \\
BB & r
\end{cases}
$$

$$
B \xrightarrow{\Gamma} C
$$

$$
C \xrightarrow{\mu} \varnothing,
$$

(1.1)
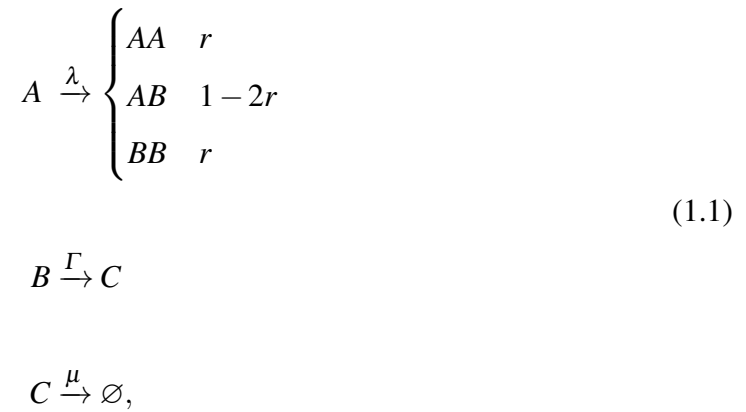
where A represents the basal layer progenitor cells, B the basal cells committed to differentiate and C the suprabasal layer cells. Progenitor cells (A) divide regularly with an overall division rate λ and give rise to either two progenitor daughters (AA), two differentiating daughters (BB) or one daughter of each type (AB) with fixed probabilities. Given the fact that AA symmetric division leads to clone expansion and BB symmetric division leads to clone extinction, the two symmetric division rates should be equal in order for a steady state in terms of number of cells to be maintained across the progenitor clone population. The probabilities of symmetric and asymmetric divisions are r and $1 - 2r$ respectively with

$0 < r < 0.5$. Differentiating daughters in the basal layer stratify to the suprabasal layer at rate Γ. Supra basal cells, C, are shed at rate μ. The proportion of progenitor cells in the basal layer is ρ and differentiating cells before stratification $1 - \rho$, (Doupé et al., 2012). As the rules of homeostasis dictate that the total basal layer cell population stay constant over time, an additional rule arises, as shown by Clayton et al. (2007), Doupé et al. (2012):

$$\Gamma = \frac{\rho}{1 - \rho}\lambda \tag{1.2}$$

Taken together, clonal dynamics in oesophageal epithelium can be fully parameterized by the division rate (λ), the stratification rate (Γ) and the probability of symmetric fate (r).

The single progenitor model of epithelial homeostasis gives rise to a number of characteristic quantitative features observed in stem cell clone size distributions. These features are considered the SP hallmarks (Clayton et al., 2007, Klein and Simons, 2011):

- The total proportion of cells remains unaltered. (Figure 1.4a)

- The average clone size of the surviving clones increases linearly with time with slope $\tau \sim \frac{r\lambda}{\rho}$ (Figure 1.4b).

- The stochastic division outcomes which include symmetric differentiation, may lead to a continuous extinction of clones. Due to this, the number of surviving clones progressively decreases, following $\frac{1}{1+\lambda rt}$ (Figure 1.4c).

- The distribution of clone sizes becomes substantially broader with time (Figure 1.4d), adopting a characteristic scaling behaviour. According to this feature, the chance of finding a clone of a size larger than some multiple of the ensemble average becomes fixed, $n/< n(t) >$ (Figure 1.4e).

a
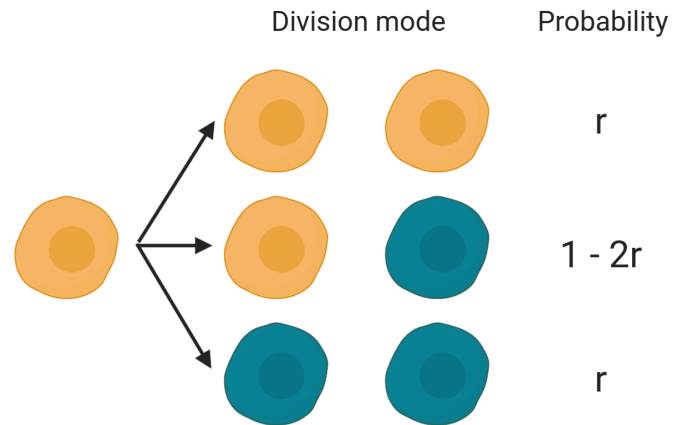
b
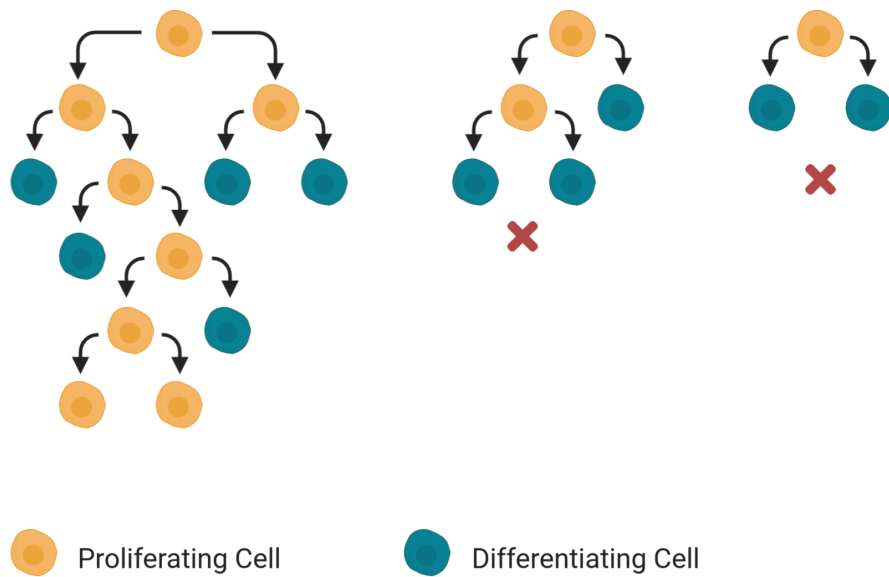
Fig. 1.3 Population asymmetry paradigm of tissue self-renewal. a) According to the SP model, stratified epithelial tissues are maintained by a single, equipotent population of progenitor cells which divide stochastically to generate either two proliferating daughters, two differentiating daughters with equal probabilities or one daughter of each type. b) Population asymmetry leads to diverse clonal populations. Due to the stochastic outcome of each division, some clones will be "lucky" and expand whereas others will inevitably become extinct as a result of symmetric differentiation.
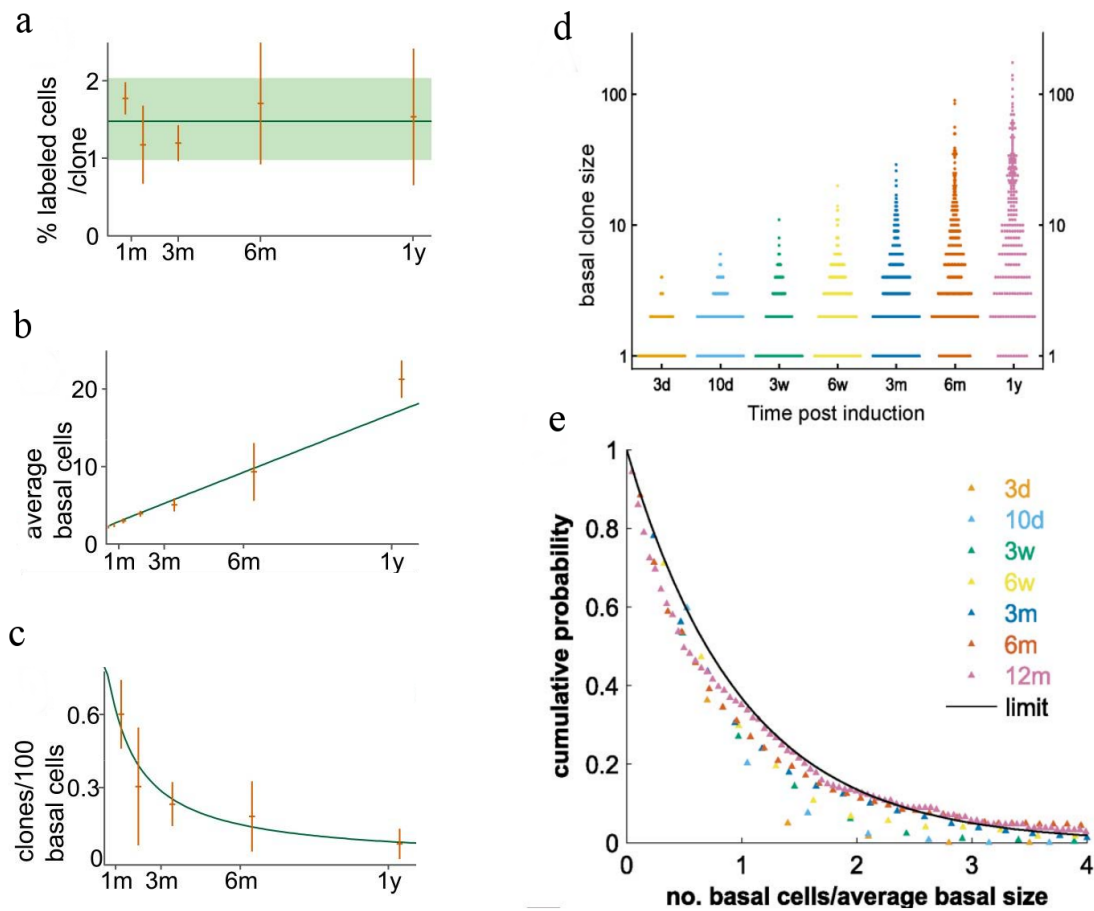
Fig. 1.4 The observation of several characteristic features in lineage tracing data supports the single progenitor as a model of epithelial homeostasis. a) the average percentage of labelled basal cells remains constant over time, b) the average clone size increases linearly with time with slope $\tau \sim \frac{r\lambda}{\rho}$, c) the number of surviving clones decreases over time following $\frac{1}{1+\lambda rt}$, d) the range of clone sizes becomes substantially broader with time, e) clone size distribution adopts a characteristic scaling behaviour so that the chance of finding a clone of a size larger than some multiple of the average becomes fixed, $n/<n(t)>$. Graphs taken from Doupé et al. (2012).

1.6   Spatial constraints determine growth dynamics

In order to study cell dynamics in epithelial tissues, one has to take into consideration that in such well defined structures, cells are organised in densely packed layers and do not act in isolation. On the contrary, their movement and growth is subject to spatial constraints imposed by their local environment and interactions with neighbouring cells.

There is growing evidence that spatiotemporal regulation of cell dynamics has important implications in both tissue maintenance and disease. Cellular interaction and competition in space are sufficient to cause alterations in growth patterns. Consistent with this, it has been recently shown that cell cycle progression of single cells in tissues may be coordinated by spatial constraints imposed by their local environment (i.e. the presence of neighbouring cells) (Streichan et al., 2014). Furthermore, analysis of mutant dynamics in healthy human epidermis suggests that the distribution of mutant clone sizes within the tissue is governed by the spatial competition of cells (Lynch et al., 2017).

Spatial constraints have been reported to have marked effects on how mutant cells expand and how tumours evolve. In support of this, it has been shown that the existence of different modes of evolution observed in human tumours can be attributed to variations in their spatial structure (Noble et al., 2019). That is to say, differences in tumour structural properties promote or limit the range of cellular interactions and spreading of mutations, thus leading to distinct modes of tumour clonal evolution.

Further evidence on the importance of space in tumour evolution comes from a recent study investigating clonal expansion in colorectal cancer xenografts (van der Heijden et al., 2019). Here, it is reported that tumour growth is driven by large clones that tend to be found at the border whereas small clones are usually located at the centre of the tumour. These findings indicate that clonal outgrowth is dictated by the spatial location of a clone within a tumour rather than cell-intrinsic factors.

Recent studies have also highlighted how disregarding the effects of space can introduce sampling bias in the analysis of mutant and cancer genomic datasets. More specifically, it has been stated that mutant clone size distributions are heavily influenced by the spatial distribution of cells within a tumour. Therefore, the underlying assumptions of how tumours grow in space in combination with the sampling procedure for sequencing tumour samples could lead to artificial biases in clonal selection (Chkhaidze et al., 2019, Noble et al., 2019) and thus misleading conclusions.

Collectively, all the above findings provide a strong evidence that space cannot be ignored when studying pre-cancer and cancer dynamics. The study of space contribution to clonal

interaction and growth within tissues is therefore key to understand the progression of mutant clones towards cancer.

1.7   Cancer in oesophageal and skin tissues

Oesophageal cancer is one of the most prevalent malignancies in the world both in incidence and mortality (Agrawal et al., 2012, Melhado et al., 2010). Oesophageal squamous cell carcinoma (OSCC) and oesophageal adenocarcinoma (OAC) are the two major subtypes of oesophageal cancer. The incidence of each subtype varies mainly according to geographical characteristics. OSCC is the most frequent type worldwide, which is present in Eastern Asia and parts of Africa, whereas OAC is a rapidly increasing disease in Western countries (Agrawal et al., 2012, Melhado et al., 2010, Qian et al., 2016, Alcolea, 2017).

There are several risk factors associated with OSCC such as diet, alcohol, tobacco chewing, smoking and physical inactivity. Other environmental influences, insufficient nutrition and drinking habits are also involved. Genetic susceptibility seems to contribute as well. The aetiology of OAC and its increasing incidence is mainly correlated with gastro-oesophageal reflux disease (GORD), Barrett's oesophagus and obesity. The overall 5-year survival rate of oesophageal cancer remains poor (15% to 25%) due to late detection (Agrawal et al., 2012, Melhado et al., 2010, Qian et al., 2016, Alcolea, 2017).

Skin cancer is distinguished in three major types: basal cell carcinoma (BCC), squamous cell carcinoma or cutaneous squamous cell carcinoma (cSCC) and melanoma. Cutaneous squamous cell carcinoma and basal cell carcinoma are the major subtypes of non-melanoma skin cancer (NMSC) and account for the 25% and 75% of NMSC respectively (Pickering et al., 2014, Madan et al., 2010). Melanoma is less common than the non-melanoma skin cancer types, but more aggressive as it is more likely to grow and metastasize (Rastrelli et al., 2014).

The primary etiological factor accounting for skin cancer is the exposure to sunlight, as it induces ultraviolet (UV) DNA damage. Considering this, it is not surprising that skin cancer occurs primarily on sun-exposed body sites. As a consequence, populations living in areas of high ambient solar irradiance have a greater risk of developing the disease. Along with environmental factors, the risk of skin cancer development is also dependent on several genotypic and phenotypic characteristics, with older and light skin colour individuals being more vulnerable (Gandini et al., 2005, Dotto and Rustgi, 2016).

In order to provide better diagnosis and therapeutic solutions for skin and oesophagus cancers, it is critical to further understand how the disease originates and evolves. To achieve

this, a better understanding of how these tissues are maintained in homeostasis, how they respond to perturbations such as wounding and how these mechanisms are deregulated during carcinogenesis should be our main focus.

1.8 The effect of Notch and p53 mutations on epithelial homeostasis

Several signalling pathways have been reported to play an important role in keeping epithelial tissues in homeostatic state as well as regulating morphogenesis during development. The Hedgehog signalling pathway has been found to regulate oesophageal epithelial basal cells' phenotypes, promoting their proliferation (van Dop et al., 2012, Rosekrans et al., 2015). It has been also shown to contribute to epidermal development, homeostasis, and repair (Abe and Tanaka, 2017). Moreover, Sox2 and p63 transcriptional regulators have been reported to be necessary for the maintenance of the tissue (Rosekrans et al., 2015, Soares and Zhou, 2018, Truong et al., 2006). The Notch signalling pathway is one of the main pathways, inducing differentiation in both skin and oesophagus. Notch is a transmembrane receptor, enabling short-range communication between cells and regulating fundamental cellular processes such as proliferation, differentiation and cell death in many tissues (Watt et al., 2008).

An increasing body of work identified several genes commonly mutated in squamous epithelial cancers. Among those, mutations in the tumour suppressor gene *TP53* and *NOTCH1* are frequently detected in squamous cell carcinoma samples (Stransky et al., 2011, Gao et al., 2014, Pickering et al., 2014, Rosekrans et al., 2015, Alcolea, 2017) Interestingly, a series of recent studies report high incidence of many of the frequently found mutated genes in tumours such as *TP53* and *NOTCH1* in healthy human skin and oesophagus (Martincorena et al., 2015, 2018, Yokoyama et al., 2019). This highlights the importance of studying the process of accumulation and interaction of these mutations within tissues in order to understand tumourigenesis.

Studying cell behaviour *in vivo* through lineage tracing offers a new avenue for understanding both homeostatic cell dynamics and more importantly changes in cellular behaviour in response to injury and key mutations that could potentially initiate and contribute to preneoplasia.

Doupé et al. studied homeostasis in mouse oesophageal epitelium and challenged the homeostatic dynamics by performing lineage tracing on wounded tissue (Doupé et al., 2012). Data showed that a single population of progenitor cells with balanced stochastic

fate maintains homeostasis and switches fate towards proliferation in response to injury conditions until the wound has healed.

Alcolea et al. studied the effects of Notch pathway inhibition in mouse oesophageal epithelial cells. To do this, the authors blocked the transcription of Notch receptor's intracellular domain using a mouse model expressing a dominant negative form of Mastermind-like 1 (Maml1) protein (DNM) (Alcolea et al., 2014). Maml1 is required in the canonical Notch pathway as it forms a complex with Notch intracellular domain, thus enabling the transcription of target genes (Kopan and Ilagan, 2009).

The dynamics of the mutated oesophageal epithelium was followed for 1 year. Over the first 3 months post induction Maml1 knock outs expanded rapidly at the expense of their wild-type neighbours. At later time points, mutant growth rate slowed down as soon as Maml1 clones started merging. However, the proportion of mutant cells kept increasing until the entire epithelium was colonised. Despite the overall epithelium being replaced by Notch mutants, no tumours were formed. Quantitative analysis of clonal data suggested that Notch mutants had their symmetric differentiation probability turned off, thus generating persistent clones (Figure 1.5a top). Interestingly, this advantage in cell fate came along with promoting the differentiation of the adjacent wild-type cells. Moreover, mutants appeared to have an increased division rate and decreased stratification rate. These observations lend Maml1 mutants an apparent competitive advantage over their wild type counterparts and explain their ability to spread across the tissue. Strikingly, once tissue take over was completed, the imbalance in mutant fate reverted towards normal, with the symmetric differentiation probability restored (Figure 1.5a bottom). A new stationary state with a faster turnover was established.

Additional treatment with carcinogens in the same experiment showed that Maml1 mutant clones could facilitate other less advantageous mutations to expand when co-existing, providing an example of the synergistic effect of different mutations in the early development of cancer.

Murai et al. studied mutant dynamics in mouse back epidermis. To this end they tracked fate behaviour of epidermal epithelial cells carrying a heterozygous p53 gain-of-function mutation $p53^{R245W}$, the mouse equivalent of the frequently detected $p53^{R248W}$ in both normal and cancerous human epidermis (Murai et al, 2018). By inducing a $p53^{R245W}$ ($p53^{*/wt}$) mutation in transgenic mouse epidermal single progenitor cells, the growth dynamics of mutants in a background of their wild-type counterparts was observed.

Up to 24 weeks post induction, the p53 mutant population outcompeted normal epidermal cells, clearly indicating their competitive advantage (Figure 1.5b top). Beyond that transient

period of increased mutant clonal growth, the expansion rate of $p53^{*/wt}$ cells weakened considerably and they eventually reached a 30% proportion of the basal layer by 15 months (Figure 1.5b bottom). Based on 5-ethynyl-2'-deoxyuridine(EdU) data, a marker that stains basal cells in S phase of the cell cycle, the restricted mutant growth at later time points could not be attributed to a decrease in mutant division rate. Along with the slowing of mutant population expansion at 24 weeks, a roughly 10% increase in basal layer cell density was observed, without influencing epidermal functional integrity. These findings suggested that $p53^{*/wt}$ epidermal progenitor cells adjust their fate to respond to alterations in their cellular environment.
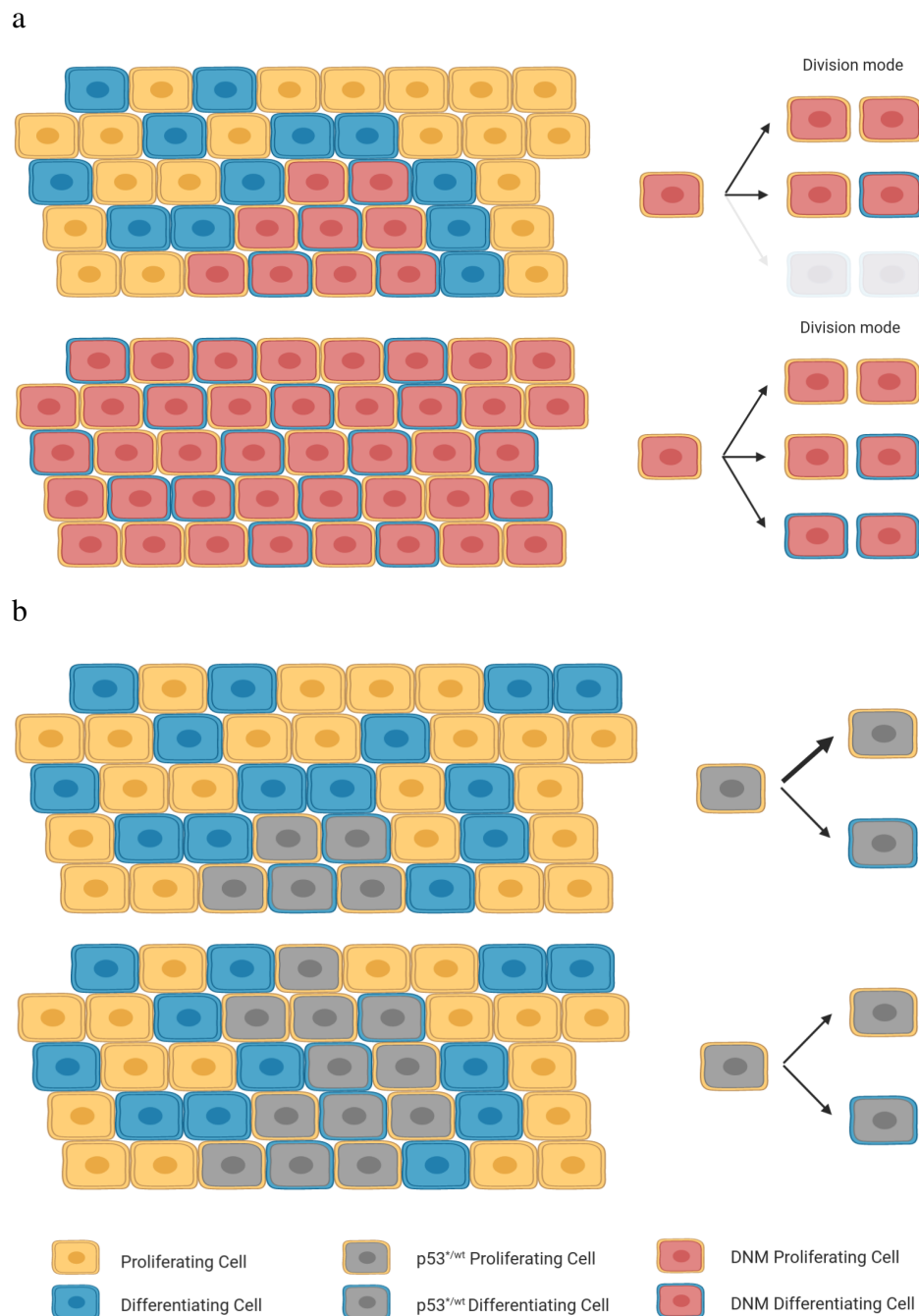
Fig. 1.5 Studying the effects of Notch pathway inhibition (a) and p53 mutations (b) in murine oesophagus and back epidermis respectively, using lineage tracing. a)Top: Initially DNM-expressing cells (Maml1 knock outs) expanded rapidly at the expense of their wild-type counterparts. The division outcome generating two differentiated cells was lost in Maml1 knock outs. Bottom: Once the tissue was colonised by DNM-expressing cells, the symmetric differentiation probability was restored and a new steady state was established. Figure based on Alcolea et al. (2014). b) Top: $p53^{*/wt}$ cells initially expanded in the tissue as they presented a competitive advantage which favoured divisions producing stem cells. Bottom: From 24 weeks post induction, the expansion of 53 mutants slowed substantially and they reached a 30% proportion of the basal layer by 15 months. The observed behaviour of mutant cells implied that crowding induced stress in the basal layer as a result of their expansion activity may have forced the return towards balanced dynamics. Figure based on Murai et al. (2018).

1.9 Modelling epithelial dynamics

Computational models can be used to give answers regarding complex biological processes, such as epithelial tissue dynamics. A variety of modelling approaches have been proposed to model epithelial maintenance, mutagenesis and cancer. These may be categorized into two main classes: continuum and cell-based models. Here, I provide a brief introduction to these modelling classes.

1.9.1 Continuum models

Continuum models have been typically used to model these processes and model cell densities or populations instead of single cells. Within this model paradigm, cells are represented as a continuous variable, thus accounting for large regions of cells with spatially averaged properties (Milde et al., 2014). Continuum models are suitable to describe the large scale tissue dynamics at the millimeter to centimeter range.

However, the approach of treating the tissue as a continuum does not preserve the identity and properties of individual cells, thus rendering the incorporation of heterogeneity between cells within a population not straightforward (Osborne et al., 2017, Milde et al., 2014).

1.9.2 Cell-based models

To underpin both normal and neoplastic tissue dynamics that result from complex and heterogeneous processes, accounting for single cell behaviour may be required. Cell-based models or individual-based models allow the simulation of individual cell behaviour explicitly. In these models, cells are represented as discrete entities with assigned properties, which respond to local interaction rules, in contrast to continuum approaches where the behaviour of individual cells is discarded.

Depending on the information required to be captured, different cell-based model classes can be used, with different levels of complexity. They can be further distinguished into on and off lattice, based on whether cells are restricted to a fixed grid or they move freely in response to applied forces (Van Liedekerke et al., 2015, Osborne et al., 2017, Metzcar et al., 2019).

On-lattice models

In the on-lattice modelling approach, space is organized into a grid of fixed lattice sites. This decreases the computational cost substantially but at the same time an accurate representation of physics is compromised.

Among on-lattice modelling approaches, cellular automata (CA) is the simplest. The concept of cellular automata was initially introduced in the 1950s by John von Neumann and Stanislaw Ulam in order to examine self-reproduction in discrete systems, and have since been extended to a series of different applications. A famous CA is the Game of life (Gardner, 1970), proposed by John Conway. The Game of life introduced and made the CA idea popular to a larger audience by highlighting the ability to study complex behaviour from simple local rules. In CA, each lattice site is typically occupied by a single biological cell (Figure 1.6a). At each time step, each cell is updated according to a set of simple local rules usually accounting the status of neighbouring cells and the space occupancy. To model processes of cell movement, division and death, each cell may either move to a vacant neighboring site, free a lattice site, or place its progeny to an adjacent site. Despite their simplicity, CA models are capable of recovering the highly complex behaviour of biological systems (Deutsch, 2005). The low complexity of this model class, allows for performing multiple simulations of large cell population sizes efficiently.

A special type of CA used for the simulation of fluid flows are lattice gas CA (LGCA) models. In LGCA models velocity channels are incorporated and a single lattice site can contain multiple cells, thus allowing the simulation of large numbers of cells (Figure 1.6b). The number of velocity channels for each lattice site reflects the number of neighbouring lattice sites. Each cell is associated with its position and velocity and the motion of groups of cells through channels between individual lattice sites is recorded (Van Liedekerke et al., 2015, Metzcar et al., 2019).

A more realistic representation of cell shapes is achieved by the Cellular Potts (CP) models (Graner and Glazier, 1992). According to this model class, multiple contiguous lattice sites may be occupied by one cell (Figure 1.6c). The evolution of the system relies on a free energy minimization approach using Monte Carlo simulation algorithm. CP models are thus capable of studying mechanical properties of the system, such as cell-cell, cell-matrix adhesion and membrane tension (Jones and Chapman, 2012). However, this modelling approach is more computationally expensive compared to the above mentioned on-lattice model classes.

Off-lattice models

A more detailed exploration of mechanical processes is achieved with the off-lattice or lattice-free models. In off-lattice models, cells move continuously in space and interact based on forces exchanged between them. The precise position of cells is tracked with the help of an equation of motion. Lattice-free models are grouped into center-based models (CBMs) that focus on cell centre interactions, where cell shape is not resolved precisely and boundary-based cell models that focus on cell boundary interactions, which allow a more detailed cell shape description, thus representing complex cell shapes (Osborne et al., 2017, Metzcar et al., 2019).

In center-based models, cells are typically simulated as spheres or deformable ellipsoids - for a more realistic morphological representation - and the centre of each cell is tracked over time (Figure 1.6d). The trajectory of each cell is described in terms of adhesive, repulsive, locomotive, and drag-like forces applied between cell centers (Metzcar et al., 2019).

A center-based modelling approach which allows a better approximation of cell biomechanics is the sub-cellular element model (Figure 1.6e). Here each cell is further divided into multiple sub-cellular elements that interact based on adhesive and repulsive forces (Newman, 2005, Milde et al., 2014).

Vertex-based models are an alternative off-lattice approach, suitable for modelling tightly packed cells in epithelial tissues (Figure 1.6f). In this model class, cells are represented as polygons whose vertices move according to equations of motion, achieving a higher resolution in cell shape and complex cell packings. Mechanical properties at the sub-cellular level can be modelled. However, the greater level of detail achieved by vertex-based modelling approach comes with an additional computational cost (Fletcher et al., 2014).
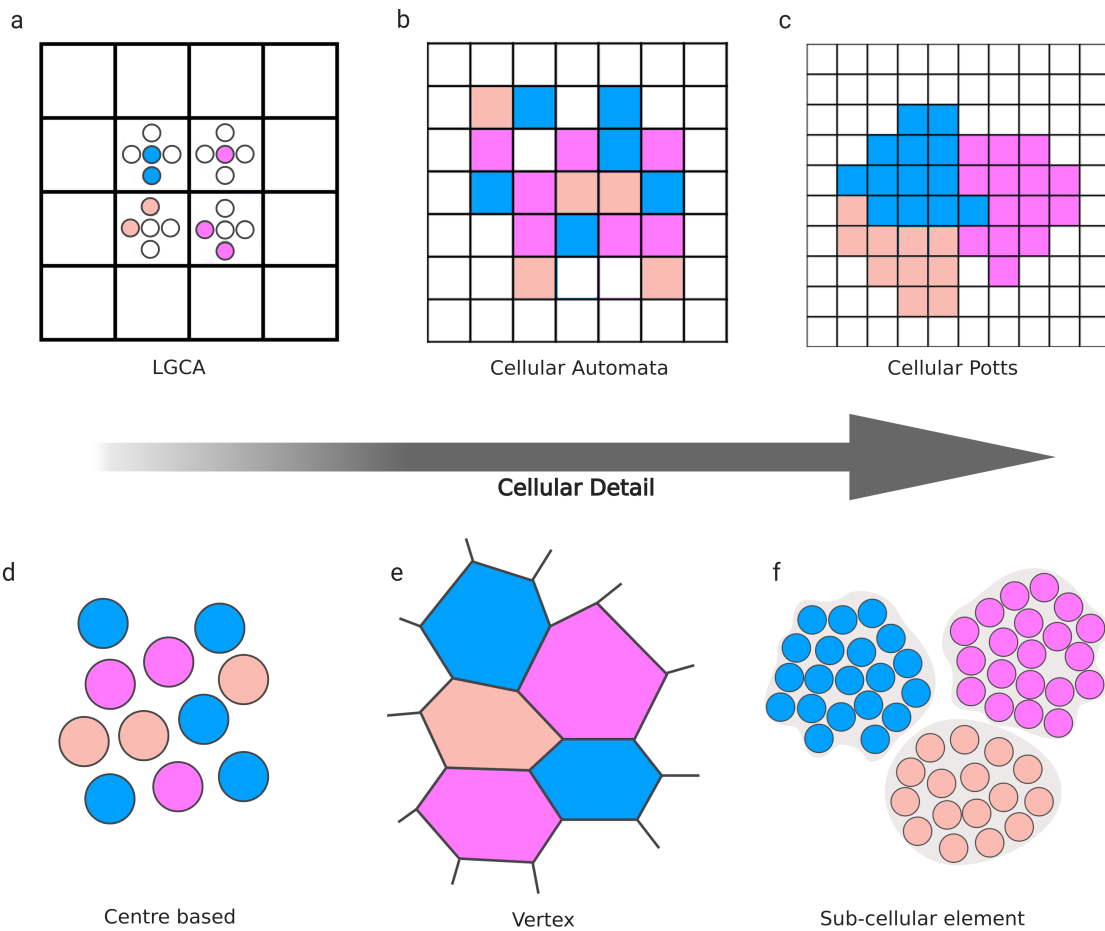
Fig. 1.6 Schematic representation of the main cell-based modelling approaches. In on-lattice models, cells are confined to a fixed grid. Lattice Gas Cellular Automata (a), Cellular Automata (b), Cellular Potts (CP) (c). In off-lattice models, cells are able to move continuously in space in response to applied forces. Centre based (d), Vertex (e), Sub-cellular element (f). Off-lattice approaches achieve higher resolution of cell mechanics but they are more computationally expensive. Figure adapted from Chamseddine and Rejniak (2019).

1.10   Aims of the project

Stratified squamous epithelia are rapidly regenerative tissues which are constantly turned over. To ensure homeostasis, the number of progenitor cells in the basal layer has to exactly match the number of cells lost from the tissue. Deregulation of these processes by mutant accumulation can result in uncontrolled chronic proliferation leading to cancer. The main focus of my research was to understand how mutations compete and spread in tissues, and how this could progress to cancer. To this end, I initially focused on analysing homeostatic

dynamics and subsequently I explored the effects of mutant and wounding events. My thesis is organised as follows:

In chapter 2, I sought to provide an explicit outline of the methods used for the purposes of this PhD project. Specifically, I presented the different approaches for computationally generating clone size distributions following the SP model. I then discussed the statistical inference techniques used for estimating model parameters. I finally described the spatial stochastic modelling approach used for simulating epithelial dynamics.

In chapter 3, I sought to challenge the underlying assumptions of the SP model to deeply assess its validity. I pinpointed issues with the proposed method for estimating model parameters and differences between experiments and model. I systematically explored different strategies for parameter inference and identified an appropriate method for accurately parameterising the homeostatic system. Accurate parameterisation will enable analysis of new datasets as well as exploration of mutagenesis.

In chapter 4, I further challenged the SP model and introduced spatial constraints to explore how these altered neutral dynamics. I implemented the SP model in the two-dimensional space in order to account for the spatial patterning of cells and gain new insights into the coordinated behaviour of cells in the context of a tightly packed tissue. The spatial model introduced transient differences in cell density across different parts of the tissue. Local extinction and crowding phenomena were observed. Importantly, the key signatures of homeostatic dynamics were successfully reproduced, suggesting that neutral dynamics remain unchanged when space is explicitly considered.

Finally, in chapter 5, I challenged the spatial SP model even further by considering non-neutral growth so that competition of advantageous mutations within the tissue could be explored. I studied non homeostatic dynamics in stratified squamous epithelial tissues by spatially modelling the growth and competition of non-neutral mutations as well as the effects of wounding in the tissue. I showed that in order to recapitulate non-neutral mutant dynamics whilst maintaining tissue turnover the spatial models need to take account of feedbacks between neighbouring cells in the tissue. Based on observed growth patterns, I suggested two distinct feedback types that could mechanistically distinguish the behaviour of p53 and Notch pathway mutations. The two distinct feedbacks mechanisms that describe mutant dynamics may suggest the distribution of mutations observed in human datasets.

Chapter 2

Methods

2.1 Generating clone size distributions

2.1.1 Markov processes

In the SP model (Section 1.5, page 8), tissue clonal evolution is described as a continuous time Markov process. This is a stochastic process $\{X(t) : t \geq 0\}$, defined by the memorylessness property (also known as the Markov property), which means that the future probability of an event is dependent solely on the current state of the system. Hence, the state at time t depends on the state at the most recent time prior to t, ignoring the states at previous time points.

If we consider a finite state space $S = \{1, ..., K\}$, a Markov process can be described as follows:

$$P(X(s+t) = j | X(s) = i, X(r) = k) = P(X(s+t) = j | X(s) = i), \qquad (2.1)$$

where times $s, t > 0$ and $r < s$. States $i, j, k \in S$.

Clonal evolution can be considered as a time homogeneous Markov process, that is to say, the way the process evolves probabilistically from a state i at time t is the same as if the process started with the state i at time 0. Thus:

$$P(X(s+t) = j | X(s) = i) = P(X(t) = j | X(0) = i) = P_{ij}(t) \qquad (2.2)$$

2.1.2 Stochastic simulations - Gillespie algorithm

Simulations of cell clonal populations following the SP model were performed based on the Gillespie algorithm or stochastic simulation algorithm (SSA) (Gillespie, 1977, Erban et al., 2007). Gillespie is a widely used stochastic algorithm for simulating the time evolution of a system and is appropriate for describing continuous Markov Processes (Banks et al., 2012).

Gillespie is an exact algorithm, where every reaction of the system is simulated, one after another, in discrete time steps. Multiple simulation runs allow the estimation of average system properties, even in complex systems where an analytical solution is not possible. From a computational point of view, Gillespie is easy to code and requires relatively small amount of memory. On the other hand, due to the fact that every reaction has to be simulated and there is need for multiple runs in order to estimate system averages, the computation time is quite long. This, in a way, imposes a limitation to the total number of reactions per run (Gillespie, 1977). In cases where the system state does not change much between time points, the speed can be increased by performing multiple reactions in a single, pre-defined time interval (tau-leaping) (Gillespie, 2001, Padgett and Ilie, 2016). Performance can also be improved by combining deterministic and stochastic approaches (hybrid models). Examples of such hybrid approaches are the Hybrid tau-leaping method, proposed by Rossinelli et al. (2008) and the Hybrid Rejection-based Stochastic Simulation Algorithm, introduced by Marchetti and colleagues (Marchetti et al., 2017, 2016).

The Gillespie algorithm was originally applied to model chemically reacting systems. The continuous increase of processing power has made the analysis of more complex systems feasible and thus Gillespie is currently used to simulate the populations and transitions of different agents in biological applications.

The following steps summarize how the algorithm was implemented to simulate the time evolution of clonal populations:

1. Initialize the number of cell types (basal proliferating cells (A), basal differentiating cells (B)) and define the event parameters (λ, r, ρ).

2. Generate two random numbers r_1, r_2 uniformly distributed in (0,1) that determine:

    (a) When the next event is going to occur (time step, dt).

    (b) Which event is going to take place. For this system there are two main event types, division and stratification, which occur at rates λ and Γ respectively.

3. Proceed to the next time point ($t + dt$) using the randomly generated time step.

4. Update the number of cell populations according to the randomly selected event.

5. Repeat steps 2-4 until there are no more cells or a predefined time threshold has been reached.

The time is determined from an event rate (the number of total events per time unit), calculated as the sum of the products of cell populations and their rates,

$$e = \lambda A + \Gamma B, \tag{2.3}$$

where e is the event rate. The time step is calculated as follows:

$$dt = -1 \frac{ln(r_1)}{e}, \tag{2.4}$$

Each event's probability is calculated as follows:

$$pAA = pBB = \frac{r\lambda A}{e}, \quad pAB = \frac{(1-2r)\lambda A}{e}, \quad pS = \frac{B\Gamma}{e}, \tag{2.5}$$

where pAA, pAB, pBB are the probabilities for the three division types and pS is the stratification probability. The next event is computed as follows:

$$A \rightarrow \begin{cases} AA, & \text{if } 0 \leq r_2 < pAA \\ AB, & \text{if } pAA \leq r_2 < pAA + pAB \\ BB, & \text{if } pAA + pAB \leq r_2 < pAA + pAB + pBB \\ pS, & \text{if } pAA + pAB + pBB \leq 1 \end{cases} \tag{2.6}$$

2.1.3 Analytical approach

Antal and Krapivsky (2010) attempted to obtain an exact analytic solution of the branching process describing the SP model ((1.1), page 9), which has recently been extended to test alternative scenarios by Greulich and Simons (2016), Parigini and Greulich (2020). As the SP model obeys a continuous time Markov process, the time evolution of proliferating (A) and differentiating (B) cell populations can be formulated in terms of the stochastic Master equation:

$$\frac{dP_{n_A,n_B}}{dt} = r(n_A - 1)P_{n_A-1,n_B} + (1 - 2r)n_A P_{n_A,n_B-1} + r(n_A + 1)P_{n_A+1,n_B-2}$$
$$+ \Gamma(n_B + 1)P_{n_A,n_B+1} - (n_A + \Gamma n_B)P_{n_A,n_B} \tag{2.7}$$

where P_{n_A,n_B} denotes the probability of finding clones containing n_A proliferating cells and n_B differentiated cells.

Antal and Krapivsky (2010) sought to provide an exact formula for calculating clone size probabilities, $P_{n_A,n_B}(t)$, and clone survival probabilities. In particular, the Master equation was solved exactly for its generating function. A generating function is a power series representation and is a convenient way to mathematically describe the branching process. Once the solution of the generating function has been found, the clone size probability distribution can be obtained from it. The generating function of $P_{n_A,n_B}(t)$ was defined as:

$$F(x,y,t) = \sum_{n_A,n_B=0}^{\infty} x^{n_A} y^{n_B} P_{n_A,n_B}(t) \tag{2.8}$$

The exact solution was obtained as shown in (2.9) and the full derivation is given in Antal and Krapivsky (2010). The equation was rewritten in terms of u, v, g and w for the purpose of simplification (2.9).

$$F = 1 - u + \frac{u(1+v) - \gamma(1+2w)}{2r} + \frac{\gamma}{2r}\frac{(1+2w)M_{1+w,0}(g) - 2CW_{1+w,0}(g)}{M_{w,0}(g) + CW_{w,0(g)}},$$
$$\text{where } g = \frac{uv}{\gamma}, \quad v = \sqrt{1-4r}, \quad w = \frac{\gamma(1-2r) - 2r}{2\gamma v} \quad u = (1-y)e^{-\gamma t}. \tag{2.9}$$

C is a constant that was determined as shown in (2.10):

$$C = \frac{-\theta M_{w,0}(\hat{g}) + (1+2w)M_{1+w,0}(\hat{g})}{\theta W_{w,0}(\hat{g}) + 2W_{1+w,0}(\hat{g})},$$
$$\text{where } \theta = 1 + 2w - \hat{g} + \frac{2r(x-y) + y - 1}{\gamma}, \quad \hat{g} = \frac{(1-y)v}{\gamma}. \tag{2.10}$$

The proposed analytical solution (2.9) was implemented in MATLAB and used by Lim et al. (2013). In particular, for every r, ρ, λ combination ($0 < r < 0.5$ and $0 < \rho < 1$) and for every time point (2.9) was called to compute the exact clone size probability distribution for a selected clone size range. This range was determined based on the average and maximum

clone size observed in experiments at that given time point. Also, in this thesis, (2.9) was re-implemented in Python and was used in the same way as in the MATLAB code. The analytical formula relied upon the use of confluent hypergeometric functions (Whittaker functions denoted by *W* and *M* in (2.9)), which were evaluated using the appropriate libraries in MATLAB (Symbolic Math Toolbox (The MathWorks, 2019)) and Python (mpmath (Johansson et al., 2013)). However, it has been shown that the computation of these functions is not trivial and remains a research topic in itself (Pearson et al., 2017). As discussed in the next chapter, existing implementations based on the calculation of these functions are time consuming and prone to errors.

2.2 Single progenitor parameter inference

To fit the SP model simulations against clonal data sets and identify appropriate parameter values, a series of different inference techniques were tested. Given that λ was measured independently from H2BGFP dilution assays (Blanpain and Simons, 2013), its value was known and the fitting was performed on parameters r and ρ.

2.2.1 Maximum likelihood estimation (MLE)

MLE is a method of estimating the parameters of a probability distribution by maximizing a likelihood function, so that model parameters with the maximum likelihood are most likely to reproduce the observed data and thus are chosen as the best fit. The likelihood $L(\theta)$ of the SP model parameters $\theta = (r, \rho)$ is the probability that the experimental data D is reproduced by the given set of parameters θ ($L(\theta) = P(D|\theta)$).

To calculate the likelihoods for the SP model parameters, a grid search was performed on a range of valid parameter values ($0 < r < 0.5$ and $0 < \rho < 1$) and the theoretical estimates of basal clone size distributions - obtained either analytically or by performing stochastic simulations - were contrasted with the ones observed experimentally by assessing the log-likelihood of every parameter combination θ. The most probable parameter combination was then selected as the parameter set with the maximum log-likelihood,

$$l(\theta; x) = \sum_t \sum_n (x_n(t) * log p_n(t, \theta)), \tag{2.11}$$

where $x_n(t)$ corresponds to the frequency of measured clone sizes with *n* basal cells at time *t*

and $p_n(t, \theta)$ is the probability of observing clones of size n at time t for a given parameter set values θ.

2.2.2 Approximate Bayesian computation (ABC)

The likelihood function lies at the core of parameter inference. However, in many applications, it is either impossible or computationally expensive to calculate the parameter space likelihood. Approximate Bayesian computation (ABC) is an inference method based on Bayesian statistics that can be used to estimate the posterior distributions of model parameters. ABC bypasses the evaluation of the likelihood function, therefore it is known as a likelihood-free method. The likelihood function is approximated by simulations, the outcomes of which are compared with the observed data. ABC methods aim - instead of finding point estimates of model parameters - to obtain the posterior distributions for those parameters (Turner and Zandt, 2012, Sunnåker et al., 2013).

The posterior distribution of a particular parameter value θ given data D can be calculated according to Bayes' theorem as follows:

$$p(\theta|D) = \frac{p(D|\theta)\pi(\theta)}{p(D)},\tag{2.12}$$

where $p(\theta|D)$ corresponds to the posterior, $p(D|\theta)$ the likelihood, $\pi(\theta)$ the prior and $p(D)$ the model evidence (i.e. the probability of the observed data given the model, also referred to as the marginal likelihood).

Approximate Bayesian computation rejection algorithm

The most basic form of ABC is the ABC rejection sampling algorithm (Pritchard et al., 1999), which is specified by the following steps:

1. A set of candidate parameter values are initially sampled from a prior distribution $\pi(\theta)$, usually chosen among a set of well known families of distributions.

2. A sampled parameter point θ^* is then used to simulate a data set D from a model specified by θ^*.

3. The simulated data X is then compared to the observed data D by calculating a distance function $d(X, D)$. If the calculated distance is less than an accepted tolerance threshold

ε $(d(X,D) \leq \varepsilon)$, θ^* is retained, otherwise θ^* is discarded. The accepted tolerance ($\varepsilon > 0$) is defined by the user.

The choice of ε lies at the intersection of computational efficiency and accuracy. Larger ε values would lead to wide distributions of samples, thus decreasing accuracy, whereas smaller ε values result in sampling very few points from the posterior distribution, increasing computational time. Typically the value of ε is chosen on a case-by-case basis such that it should be a small value but it should also achieve a reasonable acceptance rate. However, several approaches have been proposed for the choice of an appropriate ε value (Prangle et al., 2014, Faisal et al., 2013, Blum, 2010, Beaumont et al., 2002).

4. Go to step one unless there are no more candidate parameter values left or a predefined number of accepted values has been reached.

Distance based on summary statistics can be used in cases where defining an appropriate distance function between the datasets is not possible (Beaumont, 2010).

Sequential Monte Carlo ABC (SMC-ABC)

Sampling parameter values solely from the prior distribution $\pi(\theta)$ when proposing a new θ ignores both the previously accepted sampled values and the data, thus becoming inefficient. Moreover, in cases where the prior and posterior distributions are significantly different, the acceptance rate can be quite low. Sequential Monte Carlo ABC (Sisson et al., 2007) overcomes this issue by picking parameters in high posterior probability regions. Here, while iterating over successive populations, parameter points are resampled and perturbed from a weighted set of points already drawn and the tolerance ε is reduced. Perturbation allows the better exploration of the parameter space. Parameter values with distances greater than the previous ε are rejected, thus aiming to get closer to the posterior distribution on each population iteration of the sequential Monte Carlo sampler (Beaumont, 2010, Toni et al., 2009, Toni and Stumpf, 2009, Filippi et al., 2013).

The algorithm proceeds as follows (as proposed by Toni et al. (2009)):

1. Initialize $\varepsilon_1, ..., \varepsilon_T$. Set the population indicator $t = 0$.

2. Set the particle indicator $i = 1$, where particles are the sampled parameters.

   (a) If $t = 0$, sample θ^{**} independently from $\pi(\theta)$. Else, sample θ^* from the previous population θ_{t-1} with weights w_{t-1} and perturb the particle to obtain $\theta^{**} \sim (\theta|\theta^*)$,

where K_t is a perturbation kernel which specifies the transition probabilities from one state to the next.

If $\pi(\theta^{**} = 0)$ return to 2a.

Simulate a candidate data set $x \sim f(x|\theta^{**})$.

If $d(x, x_0) \geq \varepsilon_t$ return to 2a.

(b) Set $\theta_t^{(i)} = \theta^{**}$ and calculate the weight for particle $\theta_t^{(i)}$,

$$
w_t^{(i)} = \begin{cases} 1, & \text{if } t = 0, \\ \dfrac{\pi(\theta_t^{(i)})}{\sum\limits_{j=1}^{N} w_{t-1}^{(j)} K_t(\theta_{t-1}^{(j)}, \theta_t^{(i)})}, & \text{if } t > 0. \end{cases}
$$

If $i < N$, set $i = i + 1$, go to 2a.

3. If $t < T$, set $t = t + 1$, go to 2.

Particles sampled from the previous distribution are denoted by a single asterisk, and after perturbation these particles are denoted by a double asterisk.

To infer the parameters for the SP model an SMC-ABC approach was followed. Simulations of the SP model were performed starting from initial r and ρ values ($0 < r < 0.5$ and $0 < \rho < 1$) drawn from a uniform distribution, used as prior. For every simulation round, a distance metric was computed for every value pair based on the sum of the Kolmogorov-Smirnov (KS) test's distance summary statistic. KS is a statistical test that determines whether two samples or a sample with a reference probability distribution differ significantly. The KS test calculates the chance that two datasets are drawn from different distributions by calculating the maximum difference in empirical distribution functions. This distance is used with the number of observations to calculate a p-value.

2.3 Cellular Automata

In this thesis, in order to model the dynamics of stem cells in epithelial tissues, a cellular automaton (CA) approach was used. This modelling technique was chosen as it can capture appropriately the level of complexity required for analysing the available lineage tracing datasets. In this section, I provide the background of CA theory, following mainly Deutsch (2005).

2.3.1   Definition

A cellular automaton is a d-dimensional spatial model of a system, with two-dimensional CA being the most frequently used. In these models the region of space being modelled is subdivided into a number of lattice sites, whose properties are tracked as the simulated time progresses. CA describe a collection of objects, termed cells, where each of them evolves over time according to a set of rules, based on the states of neighbouring cells. Thus, following simple local rules, it is possible to see complex emergent behaviours. This principle appears to be consistent with the concept that the complex behaviour observed in many natural systems stems from the cooperative effect of simpler rules (Wolfram, 1984).

A cellular automaton may be described by the following characteristics (Deutsch, 2005):

- A discrete lattice L of sites (termed cells) and boundary conditions.

- Each cell is characterized by a finite set of states (ε).

- Each cell has an interaction neighbourhood (N^I) which is constituted by a finite set of adjacent cells.

- Each cell evolves through time according to a local transition function set of rules (R) that determine the dynamics of the states of the cells.

An example of a simple, two-state cellular automaton is depicted in Figure 2.1.

Lattice of cells

0	0	1	1	0	0	1
1	0	1	1	0	1	0
0	1	0	1	1	0	0
1	0	1	0	0	1	1
0	0	0	1	1	0	1
0	1	1	0	0	1	0
1	0	1	0	1	0	1

Neighbourhood

Fig. 2.1 Example of a two-dimensional cellular automaton model. Cells are placed on a square lattice. There are two state possibilities, 0 and 1 and the neighbourhood is defined by the 8 adjacent cells. Figure modified from Shiffman et al. (2012).

2.3.2 Lattice and boundary conditions

A regular lattice $L \subset \mathbb{R}^d$ consists of a set of cells (regular polygons) which are placed on a d-dimensional Euclidean space and each of them is labelled by its position $r \in L$. For any cell with coordinate r, the nearest neighbourhood $N_b(r)$ is defined as follows:

$$N_b(r) := \{r + c_i : c_i \in N_b, \quad i = 1, ..., b\},$$

where b is the number of nearest neighbours and N_b the nearest-neighbourhood template with elements $c_i \in \mathbb{R}^d$.

The simplest type of cell shape is square. However, lattices composed of triangles or hexagons can be used (Figure 2.2). These regular tessellations have been widely used to successfully investigate aggregate properties of a system. However, such ordered lattices might not be able to properly explain certain properties. The underlying lattice forces cell movement into fixed directions, thus generating artificial spatial anisotropies which might influence certain characteristics, such as clone shapes. To reduce this effect, irregular lattice approaches based on Voronoi diagrams were developed (Kansal et al., 2000, Gevertz and

Torquato, 2006, Block et al., 2007, Radszuweit et al., 2009). Additionally, algorithms for reducing lattice anisotropies have been introduced (Smith et al., 2019, Yates and Baker, 2013, Marek, 2013).
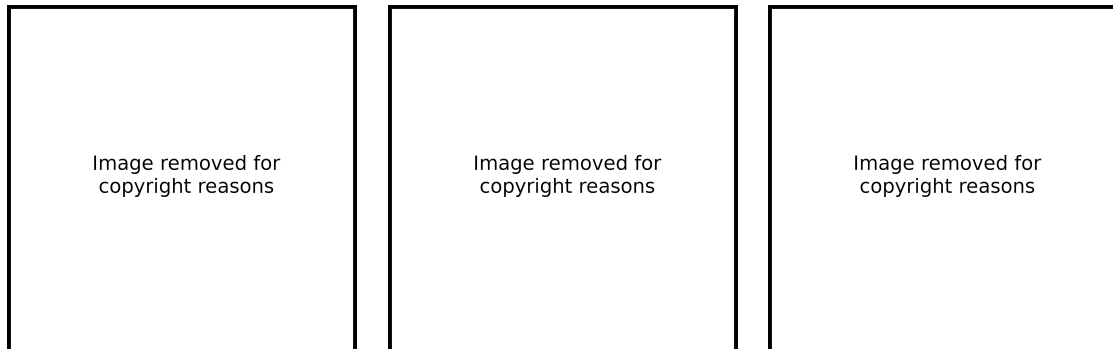
Fig. 2.2 Possible lattice types for two-dimensional cellular automata. Left: a square lattice, each cell is represented by a rectangle, middle: a hexagonal lattice, each cell is represented by a hexagon, right: triangular lattice, each cell is represented by a triangle. Figure modified from Deutsch (2005).

In CA computational experiments, the lattice is usually finite. Therefore, boundary conditions must be imposed in order to define the neighbourhood of the boundary cells. Different boundary rules may be applied. A rule often implemented for approximating an infinite lattice is periodic boundary conditions (Figure 2.3 top). In periodic boundary conditions, the opposite parts of the lattice are considered to be connected forming a ring in one dimensions and a torus in the two dimensional space. Another possible way of defining boundary conditions is to consider the lattice reflecting at each boundary (reflective boundary conditions), as shown in Figure 2.3 middle. Furthermore, the opposite parts of the lattice may be considered as a rigid wall, thus imposing fixed boundary conditions (Figure 2.3 bottom).

Image removed for copyright reasons
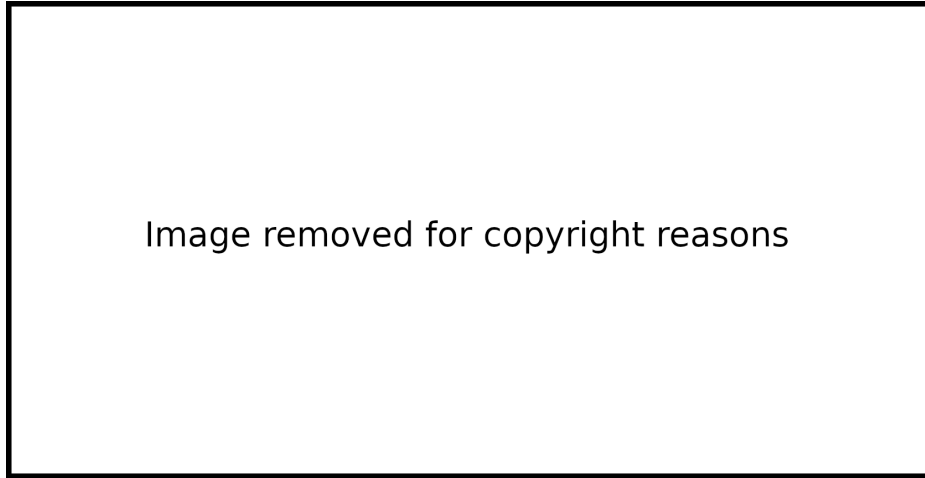
Fig. 2.3 Boundary conditions in a one-dimensional lattice, L with cells $r \in L = \{0, ..., 9\}$. Top: Periodic boundary conditions, where the opposite parts are glued together to form a ring, Middle: reflective boundary conditions, where cells at the boundaries are mirrored, Bottom: fixed boundary conditions, where lattice boundaries are considered as a rigid wall. The cells at the left and right of the dashed lines represent the nearest neighbours of the boundary cells, $r = 0$ and $r = 9$. Figure taken from Deutsch (2005).

2.3.3   Interaction neighbourhood

Each cell's behaviour in the lattice is influenced by its local neighbourhood $N_b^I(r)$, defined by a set of adjacent lattice cells. The cell itself may or may not be included in the neighbourhood. The topology of the interaction neighbourhood is specified at the beginning of the simulations and remains unchanged. Thus it is defined as an ordered set:

$$N_b^I(r) = \{r + c_i : c_i \in N_b^I, \} \subseteq L$$

Several neighbourhood topologies exist, with the von Neumann (Figure 2.4a) and Moore neighbourhood (Figure 2.4b) being the most frequently used in square lattice CA. Extended neighbourhoods may also be considered, illustrated in Figures 2.4c,d.

(a)

(b)

(c)

(d)
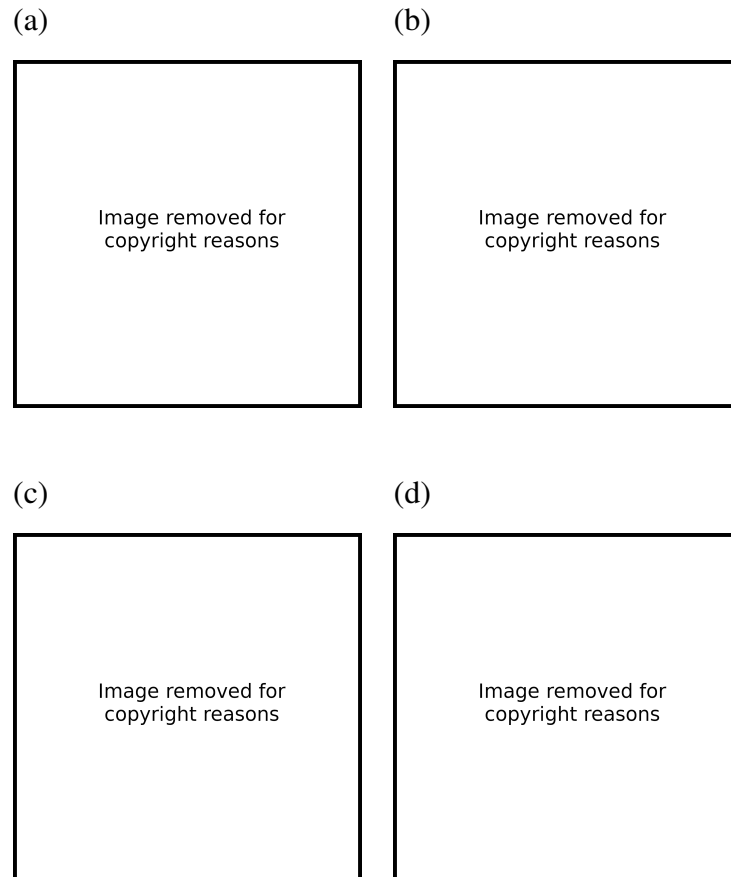
Fig. 2.4 Examples of neighbourhood types on a two-dimensional square lattice. Cell of interest is coloured in dark blue, neighbouring cells are in light blue. a) von Neumann neighbourhood, b) Moore neighbourhood, c) 2-radial neighbourhood, d) 2-axial neighbourhood. Figure modified from Deutsch (2005).

2.3.4 States

At each time step, each cell $r \in L$ is assigned a state value $s(r) \in \varepsilon$, with ε being a finite set of states. The collection of all states at a given time is called a configuration. The set of all possible state values of all lattice cells determine the global lattice configuration:

$$s := (s(r_1), ..., s(r_{|L|})) = (s(r_i))_{r_i \in L}$$

Hence, if one considers a simple one-dimensional CA with periodic boundary conditions and two possible states (on or off) for each lattice site, i.e. $s(r) \in \varepsilon = \{\text{on}, \text{off}\}$, then the lattice configuration is defined as follows:

$$\varepsilon = \{\text{on}, \text{off}\}, \quad L = \{0,..,9\}, \quad L = 10, \quad S = \{\text{on}, \text{off}\}^{10}$$

2.3.5   System dynamics

In CA, each cell in the lattice interacts with the cells that are in its local neighborhood. The temporal evolution of the system is determined by transition rules R which are executed locally on cells at discrete time steps and update the state of each cell as a function of its neighbourhood configuration:

$$R : \varepsilon^\nu \to \varepsilon, \quad \text{where } \nu = |N_b^I|$$

The local transition rules may either be deterministic (i.e. a unique next state for each cell is decided) or probabilistic / stochastic (i.e. a probability distribution of next states is specified). The system can be updated either synchronously, where all cells are assigned to their next state simultaneously, or asynchronously, where only one cell is updated per iteration.

2.3.6   Asynchronous updating

The single progenitor CA models I developed for the purposes of this thesis are probabilistic, asynchronous CA. In order to simulate the dynamics of an asynchronous CA, an appropriate updating scheme ($\mathcal{U}$) has to be introduced. If $f_u(x,t)$ represents the cell state x at time t given the updating scheme $\mathcal{U}$, the time evolution of $(x^t)_{t \in \mathbb{N}}$ can be defined as follows:

$$\forall r \in \mathbb{R}, \; x_r^{t+1} = \begin{cases} f(x_{r+n_1}^t, ..., x_{r+n_b}^t) & \text{if } r \in \mathcal{U}(t), \\ x_r^t & \text{otherwise} \end{cases}$$

The selection of the next cell to be updated was independent of the remaining cells. Every cell was assigned the time t to be updated based on an exponential distribution with mean $1/\lambda$ or $1/\Gamma$ for proliferating (A) and differentiating (B) cells respectively. The cell with the smallest t was chosen as the one to be updated in the next simulation step.

2.4   NetLogo modelling environment

The SP as well as non-neutral CA models were developed in NetLogo 6.0 (Wilensky, 1999), a free, multi-agent programming language and modelling environment for simulating the time evolution of complex phenomena in the fields of natural and social sciences. It can be used for both educational and research purposes, having a wide range of applications. NetLogo was initially created in 1999 and since then is under active maintenance and constant development. As a result of this, it is extensively documented with many helpful tutorials and also offers a large collection of example models (the Models Library), which can be used as a guidance for developing new ones.

NetLogo is written in Scala and Java and runs as a standalone application on all major computing platforms. It comes with a user friendly interface, where one has the choice of using pre-existing models found in the Models Library, modifying them according to their needs, or developing their own models. The application also supports model visualisation. This practically means that in the case of a CA model, NetLogo is able to generate and visualise the grid with boundary conditions. The NetLogo world is made up of programmable agents which interact and follow rules instructed by the model. An interface module which enables active interaction with model agents and simulations is available. NetLogo also supports command line use, which facilitates performing multiple simulation repetitions.

There is a variety of other freely available software tools for creating CA models. However, most of them are either obsolete or have not gained widespread usage. Two notable exceptions are Chaste (Mirams et al., 2013) and Repaste (North et al., 2013). Both of them are actively maintained and have been cited by several publications. The abundace of tutorials and support by the large community of NetLogo users plus its intuitive interface that allowed quick prototyping and visualization were the main reasons for selecting it.

2.5   Code

The code for implementing the non spatial SP clonal population simulations was performed in Python 3.6.5 and F# 4.0. Parameter inference scripts were written in Python 3.6.5. The existing MATLAB scripts were executed in MATLAB R2017a. CA models were developed in NetLogo 6.0. The analysis of spatial simulation outputs was written in Python 3.6.5. The code for the CA models and analysis scripts can be found at https://gitlab.com/vkostiou/SP_spatial_models.git. The runtime for a typical simulation ($100x100$ grid size, 80 weeks simulation duration) was 2.5 hours on average on a single core of an Intel(R) Xeon(R)

processor (E5-2650 v2 @ 2.60GHz). The computational complexity was $O(s)$ and $O(t)$, where s is the grid size and t is the duration of the simulation.

2.6   Graphical Figures

Graphical figures were generated using BioRender.com.

Chapter 3

Quantitative analysis of clone growth in epithelial tissues

Abstract

The stem cell dynamics of squamous epithelial cells is a major subject of study in biomedicine. The single progenitor model has been argued over several years to reproduce the observed stem cell dynamics in such tissues accurately. An analytical solution with maximum likelihood calculations has been widely used to infer the single progenitor models parameters, thus enabling access to quantitative properties of a squamous epithelial tissue in homeostasis. In this chapter, I applied the existing method to both experimental and synthetic datasets, I identified issues with the analysis, I systematically analysed the performance and compared to other inference techniques. I find that the use of the published analytical solution allows identification of a single parameter with narrow confidence intervals. However, the analysis of synthetic datasets with realistic cell-cycle distribution times and biological variation between samples suggests that these are overly precise. Furthermore, simulation-based maximum likelihood methods require extensive sampling to find a distribution of parameters making them intractable for many analyses. I conclude that an ABC based approach using a non-Markovian simulator gives appropriate error bars at an acceptable computational cost. The purposes of this analysis were two-fold. Firstly it allowed me to identify an appropriate method for analysing newly collected datasets. Secondly and most importantly, through the accurate parameterisation of the system, it provided the ability to further investigate how mutations affect tissue dynamics.

3.1   Introduction

Adult tissues rely on stem cells to establish lifelong proper function of organs. A major subject of study in biological research is the dynamics of stem cells in squamous epithelia. These are rapidly regenerative tissues covering the external surface of the body, the mouth, and the esophagus and organised in layers of keratinocytes. Importantly, most common human cancers appear in these tissues. Therefore, understanding the rules of cell fate decision is fundamental to explain not only healthy tissue growth and maintenance but also the mechanisms of wound healing, mutagenesis and cancer. A possible way to answer these types of questions would be by quantifying stem cell population dynamics.

The advent of modern genetic labelling techniques has provided useful insights on stem cell fate decision processes in squamous epithelia. Lineage tracing allows for a cell type of interest and its descendants to be genetically labelled using transgenic mice, in order to track the fate behaviour of multiple cell families (clones) at different time points (Alcolea and Jones, 2014). The distribution of cell clone sizes can then be used to further investigate clonal dynamics across the population of cells within a tissue.

There have been several studies over the past years, performing lineage tracing on epithelial tissues (Doupé et al., 2010, 2012, Mascré et al., 2012, Lim et al., 2013, Alcolea et al., 2014, Rompolas et al., 2016, Sada et al., 2016, Sanchez-Danes et al., 2016). Quantitative analysis of clone size distributions in murine oesophageal and epidermal tissues across different studies revealed that stem cell dynamics may be accurately described by a simple mathematical model, the Single Progenitor (SP) model. This supports the existence of a single, equipotent progenitor cell population which keeps proliferation and differentiation across the tissue balanced. Stem cells in the basal layer stochastically divide or differentiate through stratifying into the upper layers of the tissue before eventually being shed.

To describe tissue homeostasis, the SP model require three parameters - a division rate (λ), a stratification rate (Γ) and the probability of symmetric division (r). Using these three parameters, this model predicts average clone size, clone size distributions, tissue homeostasis, and cell survival probabilities, giving access to valuable quantitative information on phenotypic data. Estimating the model parameters accurately is critical, as it would assist in both elucidating how stem cell behave within normal tissues and more importantly how tissue dynamics change when mutations are accumulated.

3.2   Quantitative analysis of clonal data in WT tissues

3.2.1   Existing computational method for parameter inference

The progress that has been made in lineage tracing techniques over the past decades, gave rise to an increasing number of such studies. This in turn, revealed the need for appropriate solutions to analyze these datasets. To this end a computational approach has been used by previous studies (Doupé et al., 2012, Lim et al., 2013) to calculate the three SP model parameters (λ, r, ρ). This method included an analytical solution proposed in the literature (Antal and Krapivsky, 2010), which allowed for the maximum likelihood calculations. The method consisted the following steps:

1. For a range of valid ρ, λ and r values ($0 < r < 0.5$ and $0 < \rho < 1$), a probability of observing a given clone size at a given time point is calculated from an analytical solution (Antal and Krapivsky, 2010).

2. The experimental data is used to assess the log-likelihood of every parameter combination by comparing the observations (number of clones of a given size at a given time point) with the probabilities generated at the previous step. This is described in detail in Section 2.2.1, page 29. The most probable parameter combination is then selected using a maximum likelihood estimation (MLE) approach.

Initially in my analysis, I re-applied existing scripts of the most recent implementation of the above method, written in MATLAB, to infer the single progenitor model's parameters on a range of experimental datasets.

3.2.2   Experimental datasets

Clonal data of mouse oesophagus and back epidermis were included in the analysis:

- Lineage tracing of mouse oesophageal epithelium basal cells using a multicolour labelling (confetti). This technique allows the use of multicolour reporter constructs, resulting in cells randomly expressing one of four fluorescent proteins (green, GFP, cyan, CFP, yellow, YFP or red RFP). Hence, the risk of clonal merging is considerably reduced, compared to single colour systems.

  Clonal labelling was induced and the fate of labelled cells was observed at intervals of 10 days to 180 days. The number of cell clones that fall into different clone size ranges

for basal and total (basal + supra basal) cells was reported. Experiments performed by Agnieszka Wabik and Phil Jones (Wellcome Trust Sanger Institute, Hinxton, UK).

- Lineage tracing of mouse back skin basal cells (using YFP). Clonal labelling was induced and the fate of labelled cells was observed at intervals of 21 days to 545 days. The number of cell clones that fall into different clone size ranges for basal and total (basal + supra basal) cells was reported. Experiments performed by Kasumi Murai and Phil Jones (Wellcome Trust Sanger Institute, Hinxton, UK).

- In order to examine the fate of basal cells in short time scales, EdU lineage tracing data of mouse back skin cells was included in the analysis. Animals were induced with 5-ethynyl-2'-deoxyuridine (EdU) and the fate of labelled cells was tracked after 2, 4, 5, 6, 7, 8 and 10 days. The number of basal and suprabasal cells at each time point was reported. Experiments performed by Greta Skrupskelyte and Phil Jones (Wellcome Trust Sanger Institute, Hinxton, UK).

The quantitative analysis of the above experimental datasets was based on basal cell count information.

3.2.3   Oesophageal epithelium

The existing method was applied to parameterise the oesophageal experimental dataset. Considering a fixed overall division rate 1.9 ± 0.1 times/week, estimated by histone dilution experiments, a grid search for r and ρ parameters values was performed (Figure 3.1). The model predicted that the proportion of progenitor cells in the basal layer ρ is 0.7, with (0.5, 0.9 95% CI). Progenitor cells divide symmetrically to give rise to either two proliferating or two differentiating daughters with probability r=0.15, with (0.1, 0.2 95% CI). The parameter estimates seemed to fit well the experimental observations in terms of clone size distributions at all time points (Figure 3.2). Morevover, the proposed parameter values appeared consistent with the ones proposed historically by a previous quantitative analysis on mouse oesophageal clonal data ($\rho = 0.65$, $r = 0.1$), also using the same inference technique (Doupé et al., 2012).
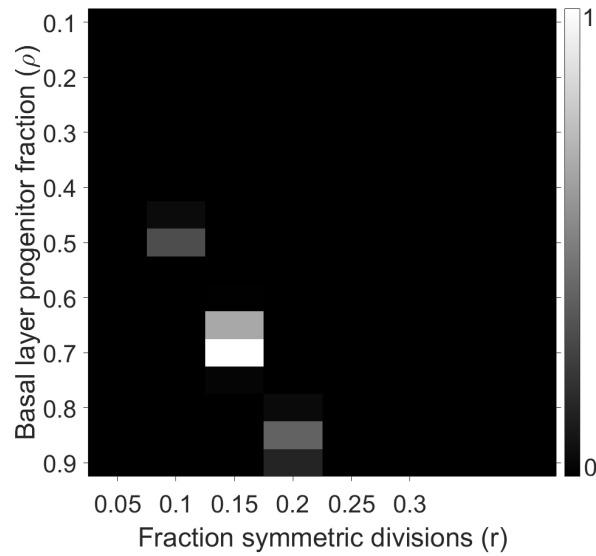
Fig. 3.1 Grid search results for the oesophagus dataset. Inference of the SP model parameters from experimental basal clone sizes, taken from the oesophageal data set. Heatmap of parameter likelihoods showing the most likely r and ρ values obtained, with prior knowledge of division rate, λ.
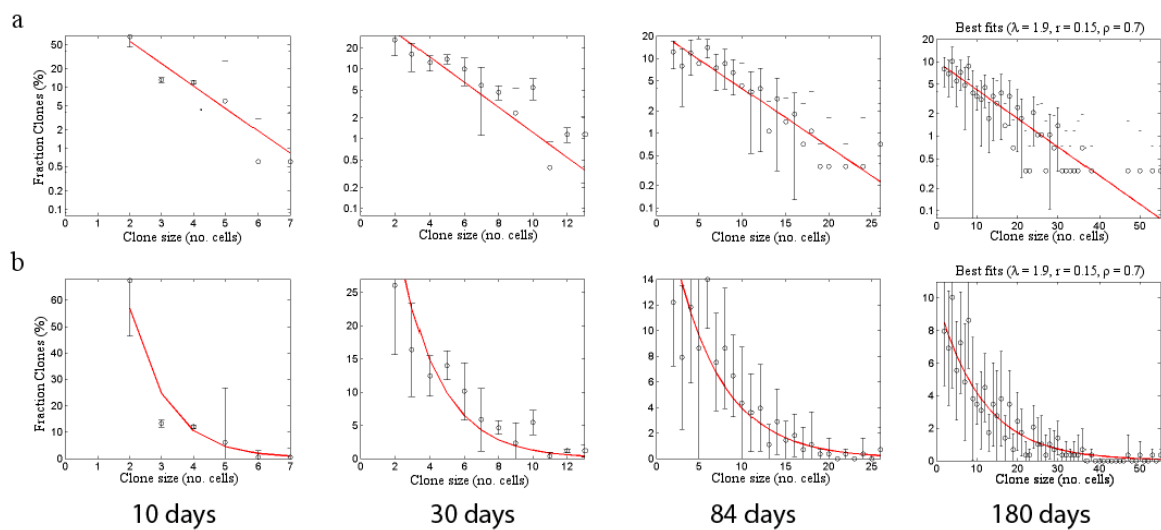
Fig. 3.2 Best fit of the model against the oesophagus dataset. Plots show best fit of the model (red line) to the experimental observations (black dots) at different time points. Data showed in a) log and b) linear scale. With a fixed overall division rate 1.9 times/week obtained by histone dilution experiments, the model predicted the proportion of progenitor cells in the basal layer ρ=0.7 and symmetric divisions probability r=0.15.

3.2.4   Back epidermis

The model was also applied to datasets of back skin, which is a different tissue that has a similar architecture to the oesophageal epithelium (Section 1.2, page 3) and thus can be considered as a comparable system. The method obtained very acceptable fits to the experimentally observed clone size distributions (Figure 3.3) and predicted the most likely r and ρ parameter combination (Figure 3.4), suggesting that the data are described by these parameters. However, the analysis failed to produce an appropriate likelihood distribution for each parameter (Figure 3.4), with likelihood distributions adopting a step-like shape. That is to say, no smooth parameter likelihood distributions were produced as several parameter combinations could not be analysed and were excluded. As a result no meaningful confidence intervals could be estimated. These observations might reflect issues in the used grid search approach. The method calculated pathological values (i.e. out of bounds probabilities) for several parameter combinations, which was mitigated by introducing small perturbations in the problematic ones. As a negative side-effect, this substantially increased runtime and did not allow using a finer grid.

Using an overall cell division rate of 1.2 times/week, calculated from histone dilution, the model predicted an r value of 0.1 and a ρ value of 0.85. As already mentioned in Section 1.5, $\Gamma = \frac{\rho}{1-\rho}\lambda$, therefore a ρ value of 0.85 would indicate a stratification rate at at 6.8 times/week which subsequently indicates that the stratification events should start at 1.029 days after division. Looking at EdU lineage tracing data in the same tissue, gave access to a more accurate estimation of ρ value. As discussed above, EdU lineage tracing provides information on cell clone dynamics in short time scales. This allows the detection of the first rounds of division and stratification. The EdU lineage tracing data clearly showed that stratification cannot take place on the first day as there were no supra basal cells detected on day 2 (Figure 3.5). On the contrary, the percentage of supra basal cells over time suggested that stratification events start after 5 days. This observation indicates that $\rho = 0.54$, which is different to the predictions of the current parameter inference engine ($\rho = 0.85$). This discrepancy in ρ might stem from differences in the experimental techniques or it might indicate an issue in the current parameter inference method.

(a)

(b)
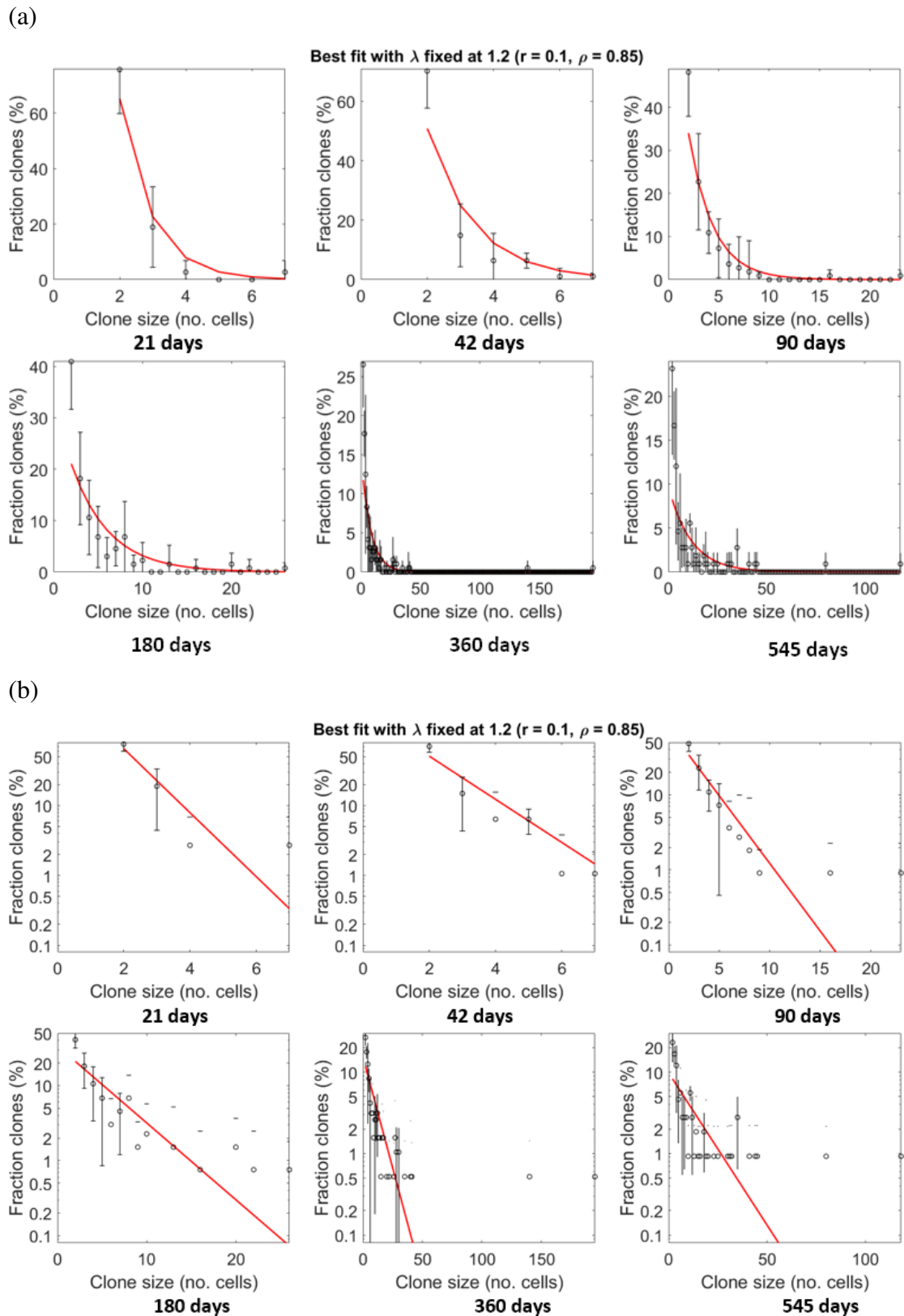
Fig. 3.3 Best fit of the model against the back epidermis data set. Plots show best fit of the model (red line) to the experimental data (black dots) at different time points. Data showed in a) log and b) linear scale. With a fixed overall division rate 1.2 times/week, obtained by histone dilution experiments, the model predicted the proportion of progenitor cells in the basal layer ρ=0.85 and symmetric divisions probability r=0.1.
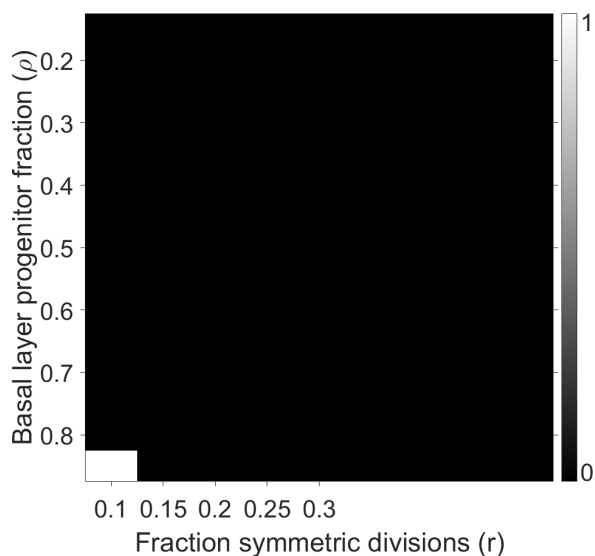
Fig. 3.4 Grid search results for the epidermal dataset. Inference of the SP model parameters from experimental basal clone sizes, taken from the back epidermis data set. Heatmap of parameter likelihoods showing the most likely r and ρ values obtained, with prior knowledge of division rate, λ. Pathological points prevented the analysis of certain parameter sets. The likelihood distributions adopted a step-like shape.
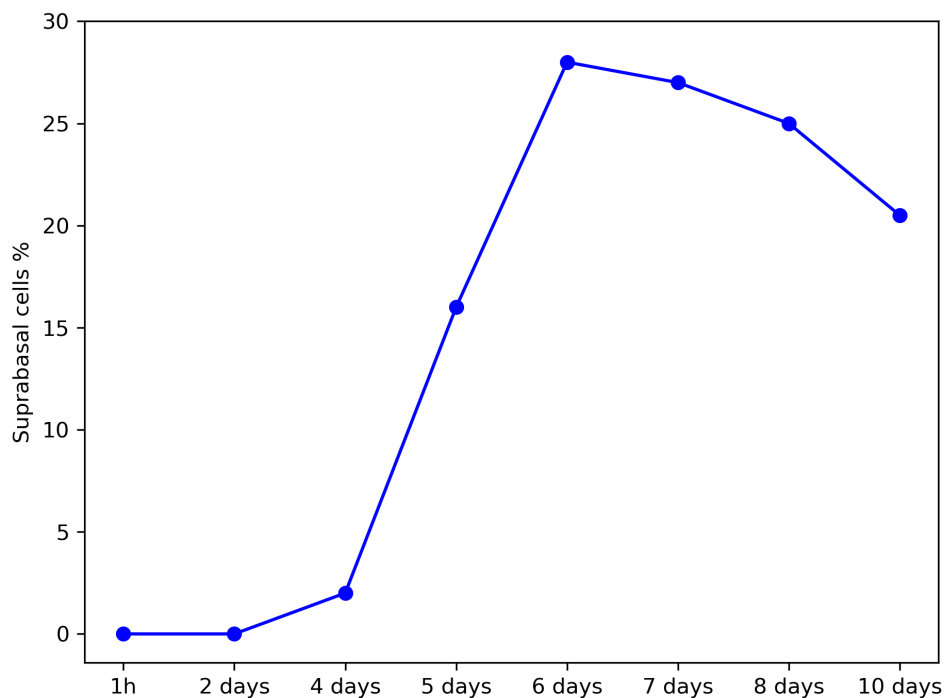
Fig. 3.5 EdU lineage tracing data of back epidermis. Percentage of supra basal cells over time. No supra basal cells were detected at 2 days. Stratification events start after 5 days, indicating a ρ value of 0.54 in contrast to the estimated ρ value of 0.85.

3.3   Synthetic data

Driven by the observed incompatibility in parameter estimation, I further investigated the robustness of the existing computational approach, by generating experimental data through simulations. This would help assessing the validity of the method on well sampled datasets with known r, ρ and λ parameters.

Computationally generated clonal datasets were produced in such a way to mimic the information contained in the experimental data, following the model. That is, the simulated clones consisted of a number of basal cells at several time points that matched the experimental ones. The artificial cell clonal simulation was based on the Gillespie algorithm (Section 2.1.2, page 26), which is a widely used stochastic algorithm for simulating how a system evolves during a given period of time.

In our studied epithelial system, described by the SP model, the cell population is maintained by basal proliferating cells A and basal differentiating cells B. Cells can divide with division rate λ, and stratify with rate Γ. There are three potential division outcomes, *AA*, *AB*, *BB*, decided by r. Table 3.1 summarizes the cell types, events and event parameters of the studied system.

Cell population types	A: Number of basal proliferating cells
	B: Number of basal cells committed to differentiate
Events	AA division
	BB division
	AB division
	Stratification
Event parameters	λ: division rate
	Γ: stratification rate
	r: probability of symmetric division

Table 3.1 Implementation of Gillespie algorithm to simulate epithelial stem cell population dynamics according to SP model. There are two cell population types, proliferating cells A and differentiating cells, B. Two main event types may take place: cells either divide or stratify. Division events are further grouped to symmetric division, *AA*, symmetric differentiation, *BB* or asymmetric division, *AB*.

Gillespie simulations of clonal populations were performed as described in Section 2.1.2. I iterated over a range of valid r, ρ and λ SP model parameter values ($0 < r < 0.5, 0 < \rho < 1$), yielding a set of 27 parameter combinations, evenly distributed in the parameter space.

Following Gillespie simulations, the synthetic datasets were generated using the following workflow:

1. From a set of 1,000,000 simulations a subset of N number was randomly picked at several time points t. N corresponds to the number of cell clones typically observed at a time point in a lineage tracing experiment and t corresponds to the time points used in such experiment.

2. The number of basal cells collected at each time point.

3. A set of size N basal cell observations from different simulations at each time point was created. A summary of the number of times each basal cell sum was observed per time point was then collected.

In studying the generated synthetic datasets, I was able to detect important features that are not observed in the real experimental datasets. Although the simulated clone size distributions were comparable to the experimental ones, a subset of simulations seemed to contain big cell clones (clones with large number of basal cell counts). These large clones were more common at late time points (t>90 days) and were observed as rare events. Figure 3.6 illustrates the largest clone size observed per timepoint across all 27 synthetic datasets.
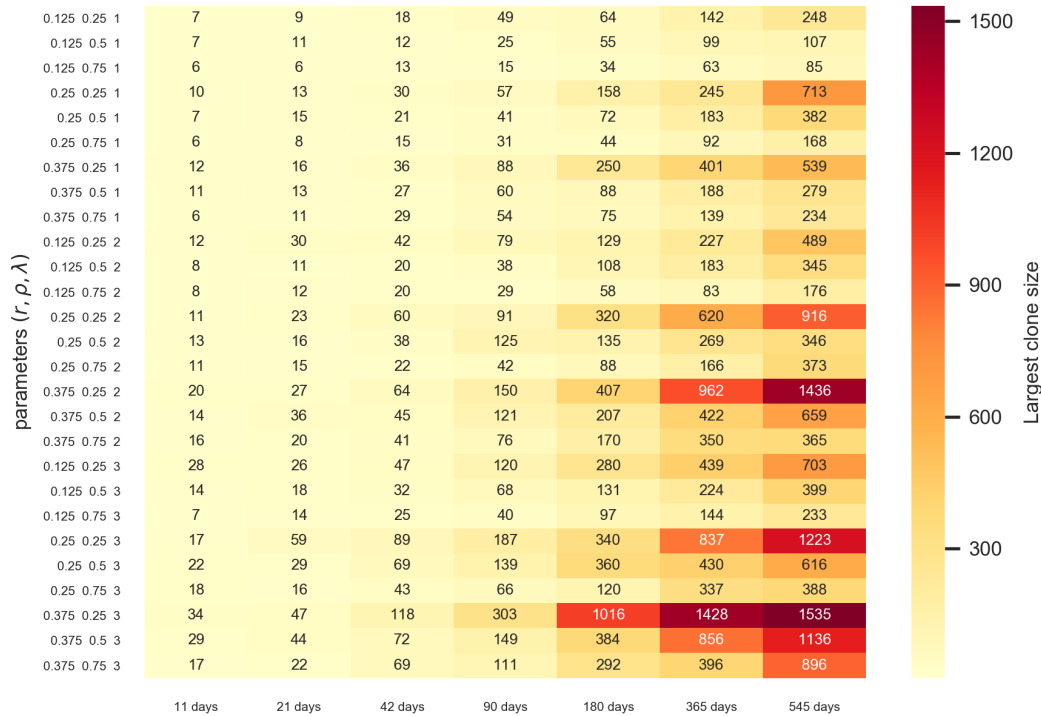
Fig. 3.6 Big sized clones across synthetic datasets. Largest clone size observed per time point for each parameter combination. Large clones were infrequently observed at late time points.

3.3.1 Validation of the simulation engine

To ensure the validity of the simulation engine both short and long timescale tests were performed. The application of these tests is critical in order to give confidence that the synthetic clonal datasets are able to follow the same behaviour as the clonal counts measured in lineage tracing experiments.

Short timescale test

The first test examined the behaviour of the computationally generated basal cell clones in short time scales.

According to the stochastic cell fate theory of the SP model, the average size of basal cell clones grow linearly over time (Clayton et al., 2007, Klein and Simons, 2011). This behaviour should be more obvious in very short time points, before stratification occurs ($t < 1/\Gamma$), and thus cell division rate (λ) is the only parameter that affects the number of

cells in the clone. Therefore, with a single cell as a starting point, the equation that describes how the number of cells increases over time is:

$$< n >= 1 + \lambda t, \quad t << 1/\Gamma \tag{3.1}$$

Considering the above, the following short timescale test was performed over all synthetic datasets: the average number of basal cells per clone should equal $1 + \lambda t$ and thus should scale linearly over time. Values of slope, intercept and r-squared of linear regression were calculated. The expected linear behaviour was successfully reproduced as shown in Figure 3.7. In particular, slope value approximated the expected λ value when $t << 1/\Gamma$.

(a)

(b)

(c)
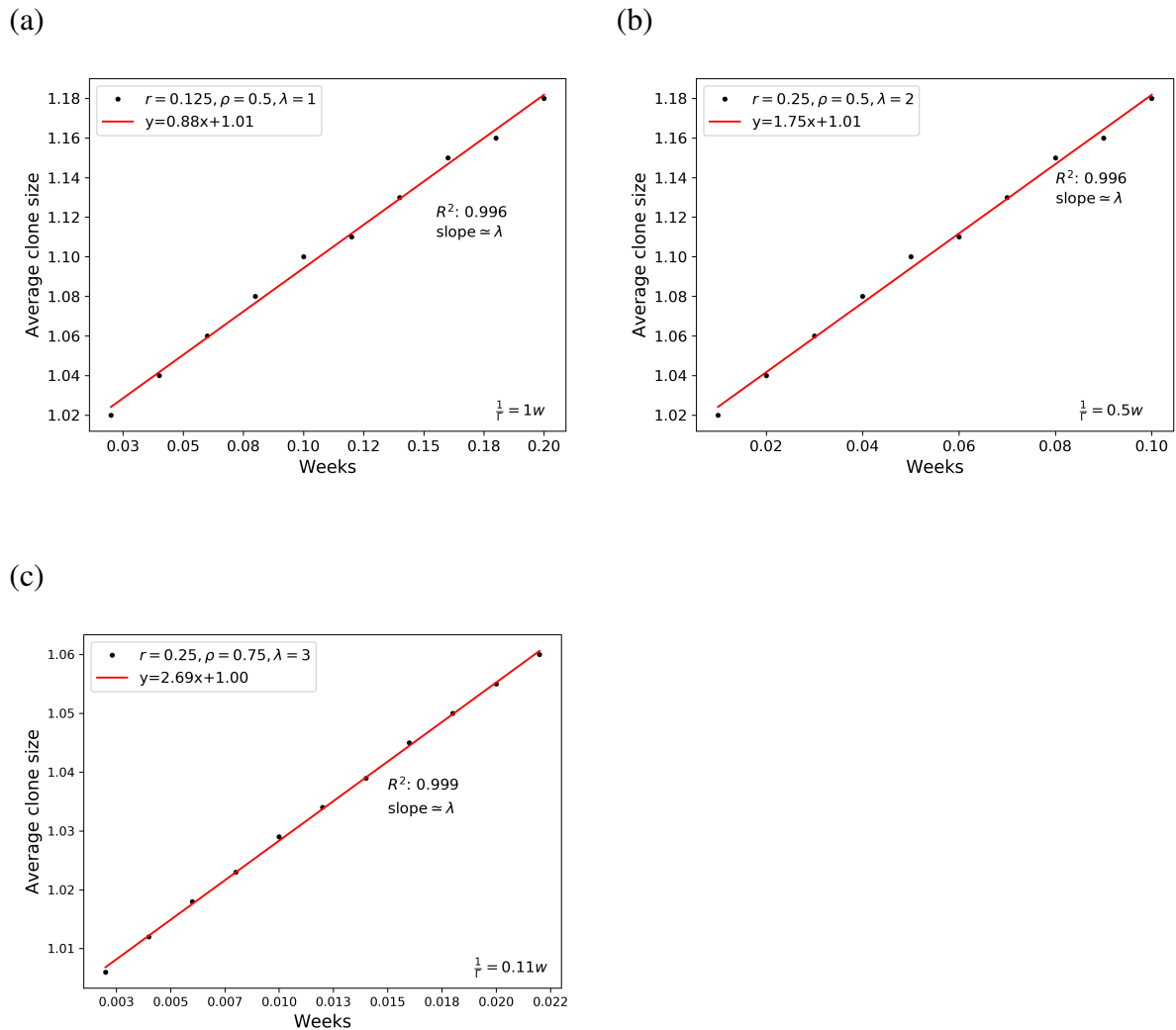
Fig. 3.7 Short timescale test. Typical average basal clone size plots at early time points, as calculated from synthetic data in order to perform the short timescale test (a,b,c): The average number of basal cells per clone scales linearly over time, before stratification occurs ($t << 1/\Gamma$), and is described by $<n> = 1 + \lambda t$ linear equation.

Long timescale test

The expected clone size behaviour at long time points according to SP model is that the average number of basal cells per clone grows linearly with time: $<n> = \frac{r\lambda}{\rho}t + \frac{1}{\rho}$ (Clayton et al. 2007). Therefore, the long timescale sanity check involved calculating the average clone size at regular time points from 4 weeks to 100 weeks to examine if they scale linearly with time. Given the fact that at long time points, basal clone size distribution is affected by stratification events and thus persisting or extinct clones can be observed, two different approaches were considered. In the first approach, cell clones with zero size were excluded

and only the surviving ones were included (Figure 3.8a,b), whereas in the second the average number of basal cells included all cell clones (surviving and extinct) (Figure 3.8c,d). After the application of the two above approaches I was able to show that the expected behaviour was accurately reproduced. In addition, the average number of basal cells including all clones remained constant, showing that the system becomes homeostatic with average clone population $\frac{1}{\rho}$ (Figure 3.8c,d).

(a)

(b)

(c)

(b)
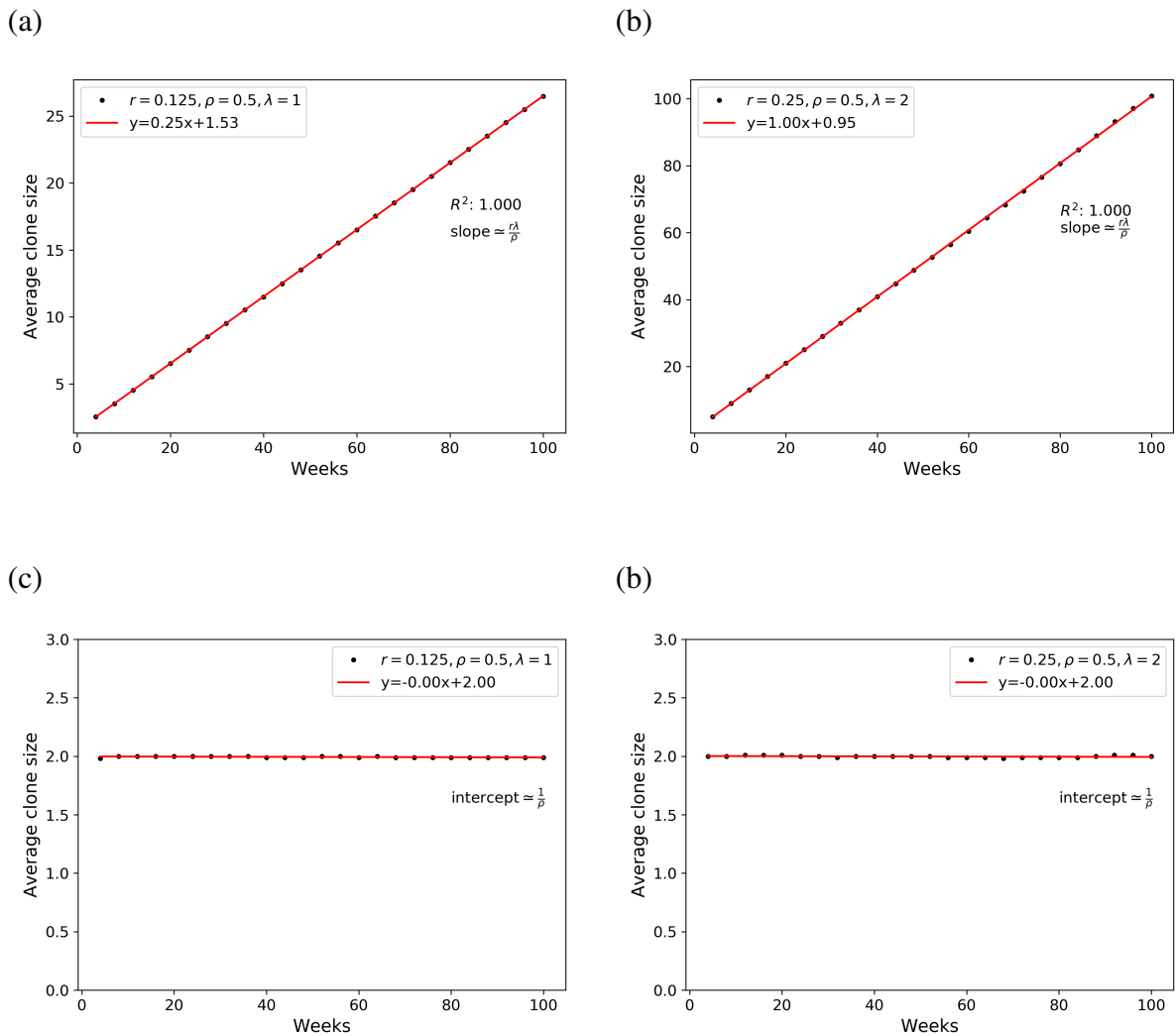
Fig. 3.8 Long timescale test. The long time scale test successfully reproduces the expected clonal behaviour. Long term average basal behaviour of survival (a,b) and survival plus extinct (c,d) clones, as calculated from synthetic data: a,b: The average number of basal cells per clone scales linearly over time with slope $\tau \sim \frac{r\lambda}{\rho}$, c,d: the average number of basal cells including all clones remained constant, showing that the system becomes homeostatic with average clone population $\frac{1}{\rho}$.

3.3.2   Quantitative analysis of synthetic data

Once the results of the sanity checks confirmed that the synthetic datasets accurately mimicked the real experimental data, I proceeded to the analysis step. Analysis was performed exclusively on basal cell counts. Each of the 27 computationally generated datasets was used as input to the existing computational method in order for the r, ρ and λ parameter likelihoods to be estimated. As λ value is usually calculated from alternative experiments, analysis was performed on the remaining r and ρ values. Given that each input synthetic dataset depicts how stem cell population is distributed at given r, ρ and λ, the method should be able to infer the values of the unknown parameters.

The results have clearly raised several issues in the published computational approach. The first problem that arose was the inability of the method to analyze large size clones, a feature observed in the synthetic datasets, especially at late time points (see Section 3.3). Observations of clones of large size were highly improbable resulting in the sum of all likelihoods being –inf. As a consequence, the probability values for observing large clones could not be represented within machine precision (64 bits), rendering the analysis of the majority of the synthetic data not feasible. Modifying the code by converting to log probabilities when large sized clones were observed allowed me to analyze the whole cohort of the synthetic datasets.

Whilst the estimated parameters appeared to achieve a very acceptable fit in clone size distributions with input data, as shown in Figure 3.9, they did not match the input values as it would be expected (Figure 3.10). The discordance between input and output r and ρ parameters was consistent across all synthetic datasets, highlighting an issue in the accuracy of parameter inference (Figure 3.11).
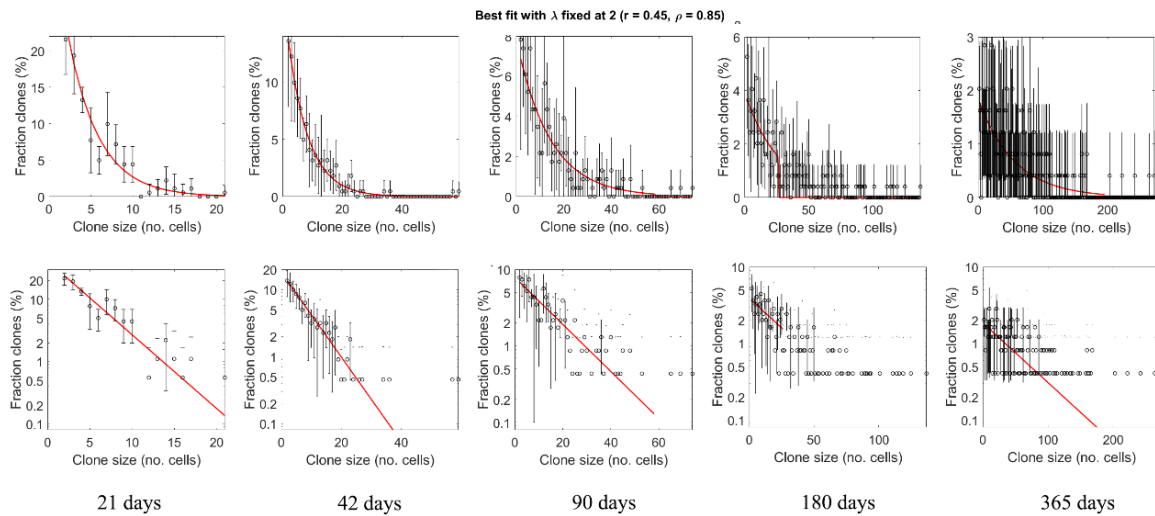
Fig. 3.9 **Best fit of a synthetic dataset against the model.** Plots show best fit of the synthetic data with input parameters $\lambda = 2$/week, $r = 0.25$, $\rho = 0.5$ (black dots) to the model (red line) at different time points. Data showed in a) linear and b) log scale. With a fixed overall division rate 2 times/week the model predicted the proportion of progenitor cells in the basal layer ρ=85% and symmetric divisions probability r=0.45.
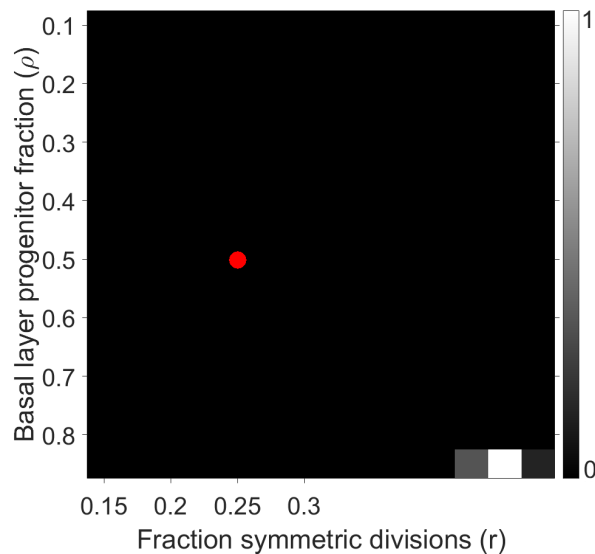
Fig. 3.10 **Grid search results for an example synthetic dataset.** Inference of the SP model parameters from basal clone sizes taken from synthetic data with input parameters $\lambda = 2$/week, $r = 0.25$, $\rho = 0.5$. Heatmap of parameter likelihoods showing the most likely r and ρ values obtained, assuming prior knowledge of division rate, λ. Red dot displays the expected r and ρ values, used to generate the synthetic dataset. The most likely inferred parameters do not match the expected ones.
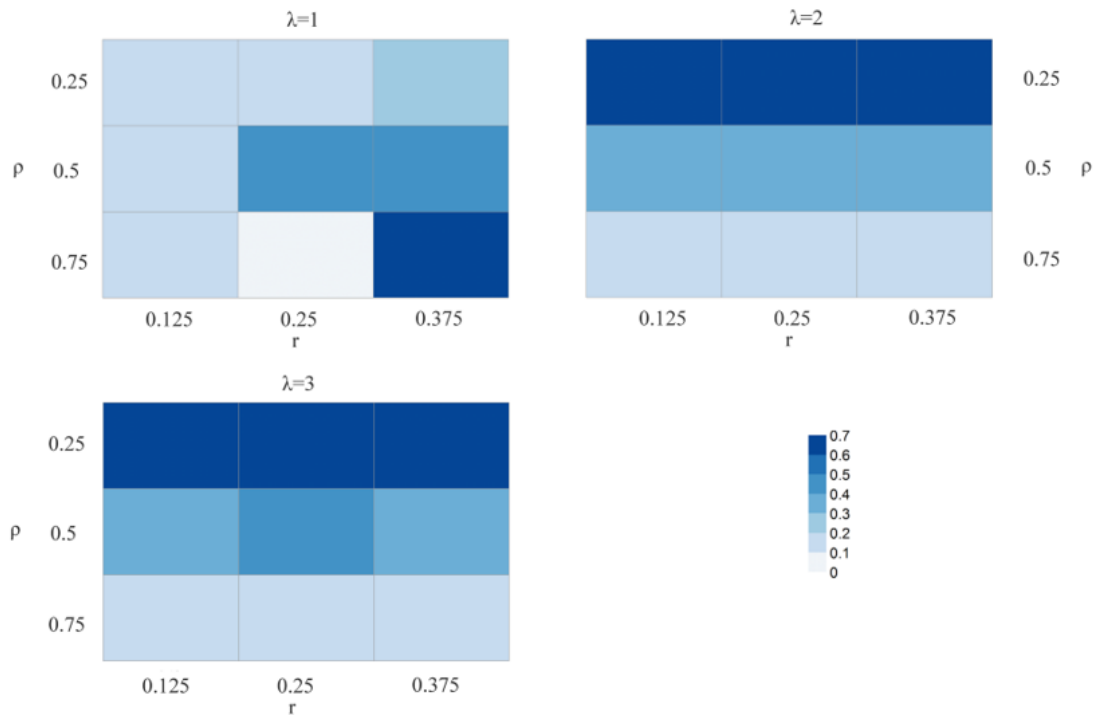
Fig. 3.11 Input and estimated parameter values for all synthetic datasets. The Euclidean distance between input and estimated parameter combinations highlights the issues in the accuracy of parameter estimation. Higher distance indicates higher discrepancy between input and output parameter values.

Synthetic replicates and increased sampling

In order to further investigate the power of the above-mentioned conclusions, and to establish whether incorrect predictions arose from poor sampling, the synthetic data analysis was repeated by generating two types of alternative synthetic datasets:

- Replicates, to show how much the outcomes can vary the same synthetic datasets with different samples.

  Synthetic replicates were produced using the same input parameters but picking a different subset of simulations. No substantial differences in the estimated parameters were observed between the replicates and the original datasets.

- Datasets with increased sampling, to find out how many clones would need to be counted in order to get accurate estimates of different parameters. This analysis was done to examine whether a potential source of the inaccurate parameter estimation is the existing sampling regime. That would be a source of experimental error that could be trivially fixed.

Synthetic data with increased sampling were produced using the simulation engine and picking a larger subset of simulations. More specifically, two datasets of 500 and 5000 cell clones were generated. Suprisingly, no improvement in the parameter estimation was observed even when the sampling regime changed by counting more cell clones.

3.4 Issues in the existing parameter estimation method

The analysis of new experimental and synthetic data sets highlighted important issues in the underlying software used for clonal data analysis that might be responsible for the inaccurate parameter estimation. An already mentioned problem is the inability to estimate the probabilities of big clone sizes without introducing modifications to the code (see Section 3.3.2, page 55). Furthermore, as also shown by the analysis of experimental clonal data, the method fails to produce smooth parameter likelihood distributions and reasonable confidence intervals.

The most important issue though was the observation of collision in estimated probabilities of different parameter inputs. Table 3.2 illustrates an example case of discordant synthetic input and estimated parameters.

Input r	Input ρ	Input λ	Output r	Output ρ	Output λ	Euclidean Distance
0.25	0.5	2	0.425	0.85	2	0.4

Table 3.2 Input and estimated parameter values for an example synthetic dataset. The Euclidean distance between input and estimated values highlights the issues in the accuracy of parameter estimation. Higher distance indicates higher discrepancy. The Euclidean distance was calculated as: $\sqrt{(r_i - r_o)^2 + (\rho_o - \rho_i)^2 + (\lambda_o - \lambda_i)^2}$, where i, o correspond to input and output respectively.

The current method was used to calculate the probabilities for observing clone sizes 2-10 at 3 and 6 weeks for both input and output parameters (Figure 3.12). The outcome of this experiment revealed that the two different parameter combinations gave the same clone size probability distribution. Given the fact that this method is based on calculating the probabilities of observing a given clone size at a given time point for a given ρ, λ and r parameter combination and compare them with the experimental data to infer the most likely parameter combination, it is highly problematic that same clone size probabilities are evaluated for different parameter values. Collectively, all the issues discovered from testing the existing computational method highlight the need for a more accurate parameter inference approach.
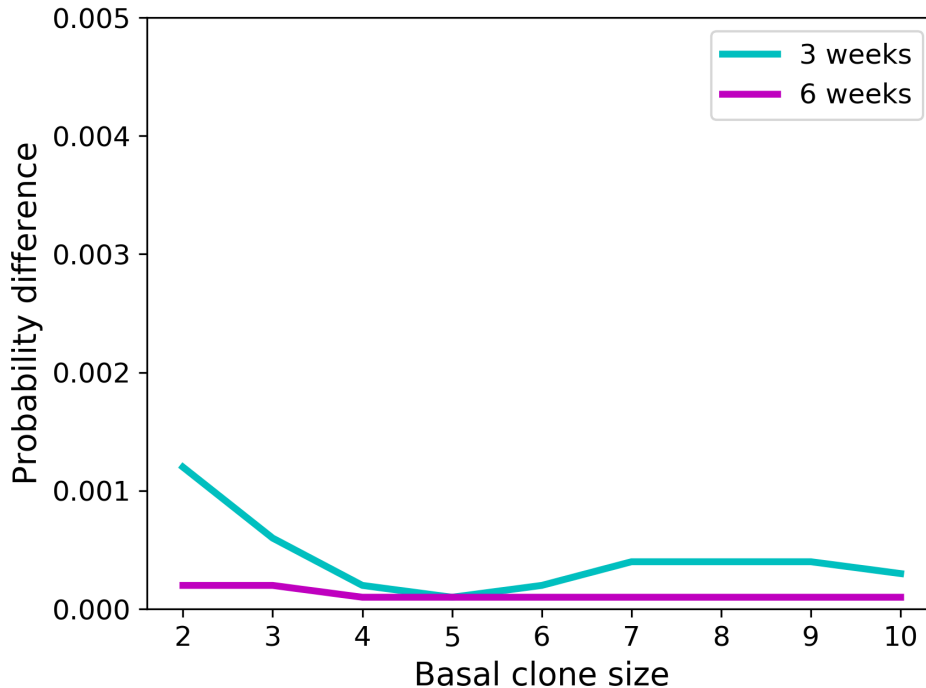
Fig. 3.12 Multiple acceptable solutions for the same dataset. The difference of clone size probability distributions between input and estimated parameters as calculated by giving the $r = 0.25$, $\rho = 0.5$, $\lambda = 2/\text{week}$ synthetic dataset. Despite the fact that the estimated parameter values did not match the input ones the difference in their probability values is almost zero, indicating same probability distributions for different parameters.

3.5 Exploring alternative approaches for clonal data quantification

In seeking to define a more robust methodology for estimating clonal data parameters, I tried several alternative approaches. The analytical solution is reliant on the computation of confluent hypergeometric functions, which is a non-trivial task and is a research topic in itself (Pearson et al., 2017). The existing MATLAB implementation for calculating these equations is time consuming and prone to error in many parameter regimes (see Section 3.2.4, page 46). Considering this, I initially re-implemented the analytical solution in Python. Estimating clone size probabilities by alternative functions might eliminate the probability collision, the most serious issue that emerged from the validation of the existing parameter inference method.

The new implementation of the analytical engine using Python libraries was applied to a set of synthetic data. Strikingly, the inferred parameters appear to match the input ones, for a

series of different synthetic datasets (Figure 3.13). Nevertheless, the method was not able to produce appropriate parameter likelihood distributions, as the calculated likelihoods were very precise with narrow confidence intervals.

Despite its success in estimating the correct parameters for the given set of synthetic data, the analytical approach has still limitations. A known issue in the analytical solution is the assumption of an exponential distribution of stem cell cycle times. This is restrictive, particularly in light of short term data from histone labelling and live imaging experiments (Rompolas et al., 2016), showing that cell cycle distribution times do not following exponential distributions (Piedrafita et al., 2020). With that in mind, a series of inference approaches which have the flexibility of accounting for more realistic cell cycle times were explored. A common alternative to analytical solutions are simulation based techniques. Considering this direction, I performed Gillespie simulations to calculate the probability of observing a clone size at a given time point. To perform an initial comparison, I considered exponentially distributed cell cycle times. The generated clone size probabilities were then contrasted with clonal data using a set of different methods.

- Maximum Likelihood estimation (Section 2.2.1, page 29): the frequency of each basal clone size per time point, obtained by Gillespie SP simulations was compared with the clonal data in order to calculate a log-likelihood value. As indicated in Section 3.3.2, SP model's characteristic scaling behaviour may lead to large variations in the clone size distributions. This would imply that if certain clone sizes are not represented in the stochastic simulations, this might introduce a bias in the log-likelihood calculations. To deal with this issue, I decided on binning the clone sizes in ranges increasing in powers of two.

- Cumulative distance function: the closeness between the SP simulations and the experimental data was estimated by computing the sum of the Kolmogorov-Smirnov (KS) test's distance statistic.

- Sequential Monte Carlo ABC (SMC-ABC) (Section 2.2.2, page 30): In SMC-ABC the initial parameter set is generated randomly and the quality of fit between model simulations and experiments is measured according to a distance function. The new parameter sets to be tested are derived from a perturbation of the previously checked ones. By rejecting parameter values with calculated distance above a certain threshold, the most likely parameter combination is estimated after certain rounds of tries.

(a)

(b)

(c)

(d)
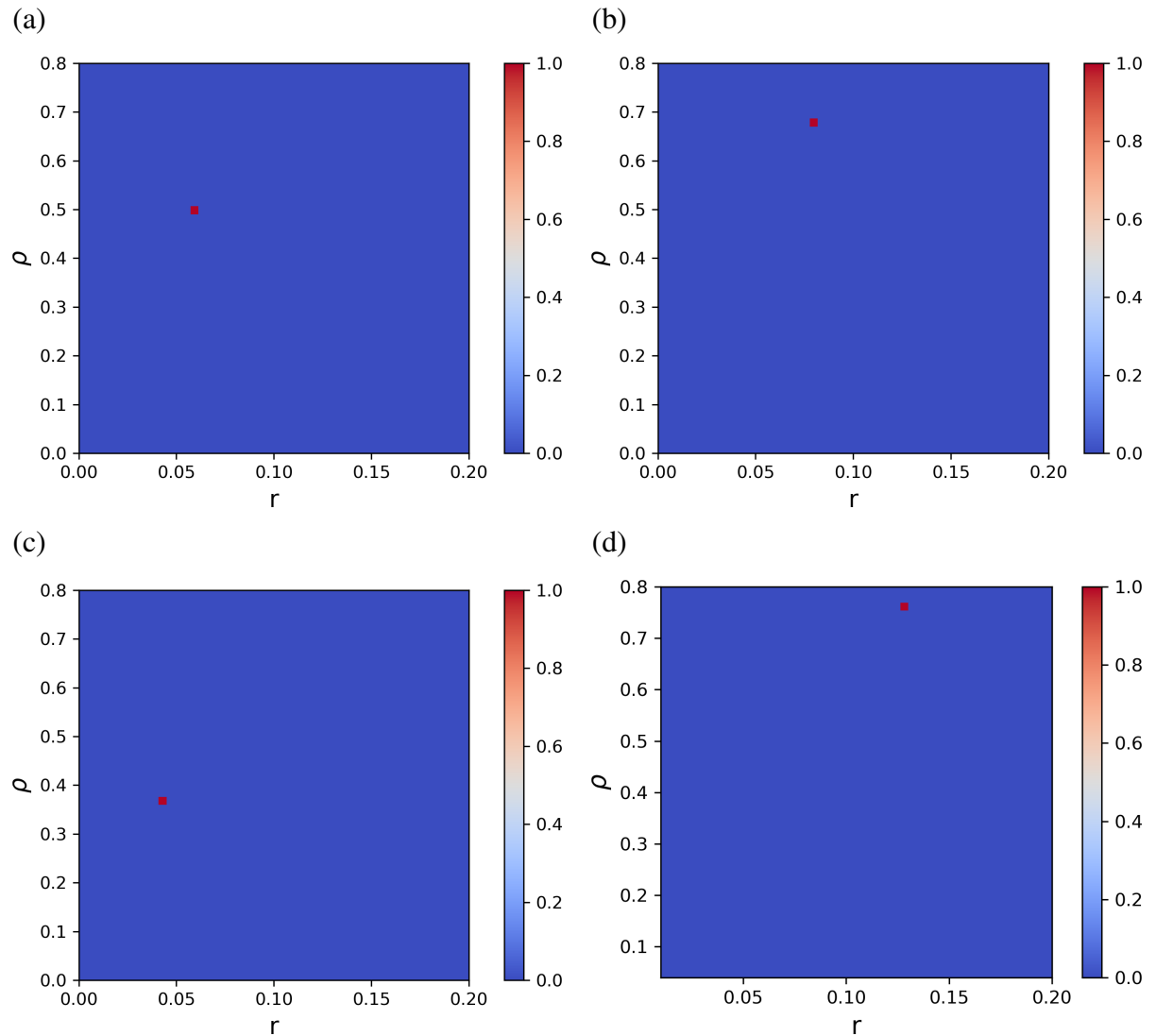
Fig. 3.13 Synthetic data analysis using the new analytical engine. The new analytical engine with a maximum likelihood calculation is able to successfully infer r and ρ parameters when applied to synthetic data with exponentially distributed cell cycle times. Heatmap of parameter likelihoods showing the most likely r and ρ values obtained, assuming prior knowledge of division rate, λ. a) $\lambda = 2.9$/week, $r = 0.06$, $\rho = 0.5$, b) $\lambda = 2.9$/week, $r = 0.08$, $\rho = 0.66$, c) $\lambda = 2.9$/week, $r = 0.045$, $\rho = 0.375$, d) $\lambda = 2$/week, $r = 0.125$, $\rho = 0.75$. Results are very precise, with no smooth parameter likelihood distributions.

To explore the limitations of the different methods and find the most appropriate strategy for parameterising clonal data, the analytical engine and the simulation based alternative approaches were applied to the analysis of both previously published experimental (Doupé et al., 2012) and synthetic datasets with exponential cell cycle distributions.

Synthetic and experimental data analysis by different methods produced similar outputs, with all approaches broadly agreeing on the estimated parameter values (Figures 3.14, 3.15).

In terms of computational efficiency, SMC-ABC and the analytical approach were the fastest methods, with the SMC-ABC also being able to take advantage of multiple CPU cores. On the contrary, simulation based approaches required more time to complete the calculations which increased substantially with sample size.

One of the hallmarks of the SP model is that the average basal clone size grows linearly with slope of this line being equal to r/ρ (see Section 1.5, page 8). Driven by this, I performed a restricted search assuming a relationship of r/ρ calculated from the slope of the average clone size in order to reduce the parameter search space significantly and allow for increased sampling. Both simulation and analytical methods agreed on the estimated parameter values (Figure 3.16). The analytical solution was able to infer the expected parameter values producing narrow confidence intervals and a smooth likelihood distribution (Figure 3.16a). On the contrary, maximum likelihood based simulations required a sample size larger than 100,000 simulations per parameter set to produce an appropriate distribution, increasing computational effort substantially (Figure 3.16b).

Collectively, all the above observations suggest the newly implemented analytical solution and the SMC-ABC as more efficient methods for parameterising the SP model.
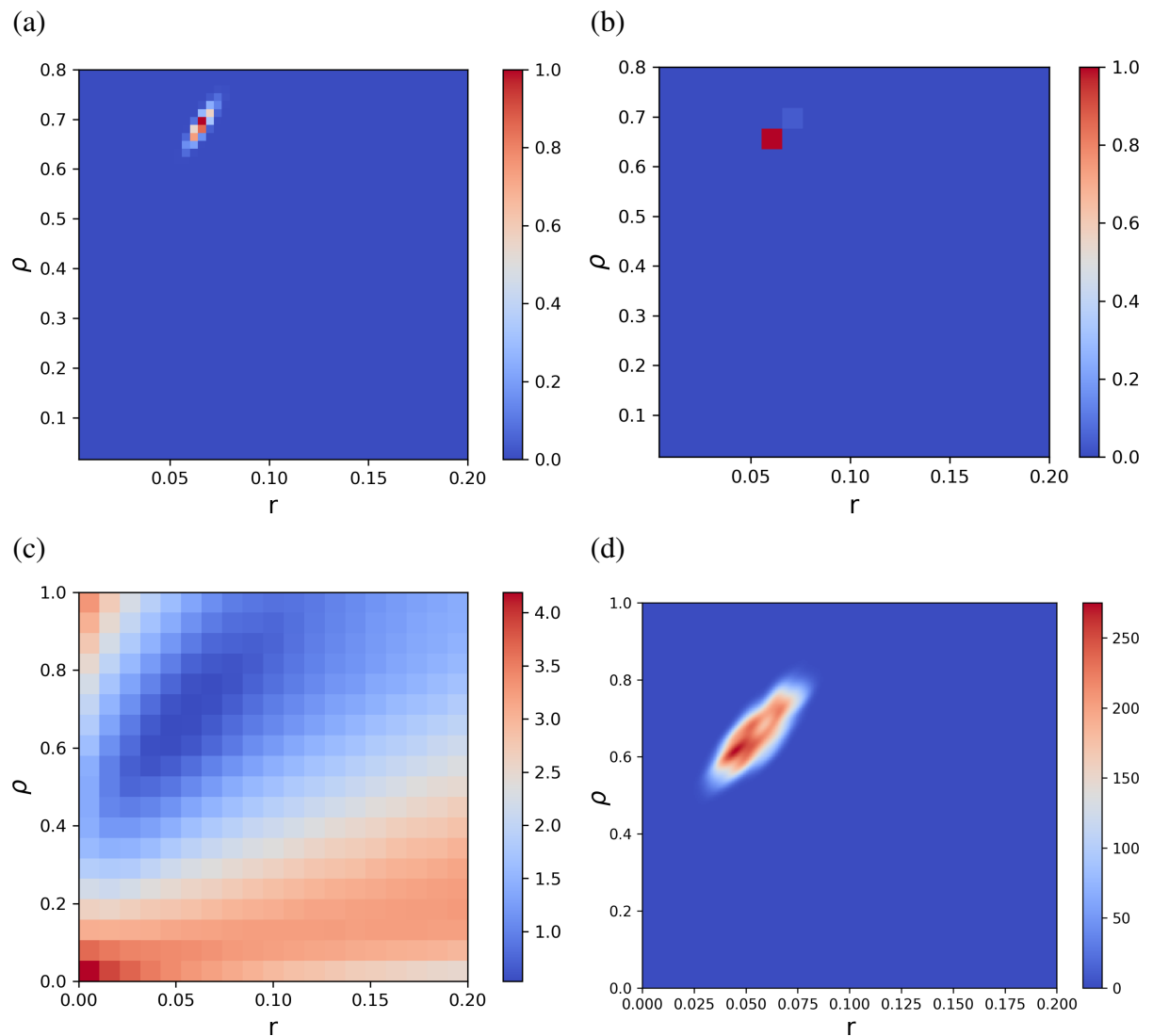
Fig. 3.14 Analytical solution and simulation-based methods attempting to infer SP parameters when applied to oesophageal clonal data from Doupé et al. (2012). Published parameters: $r = 0.1$, $\rho = 0.65$, $\lambda = 1.9$/week. Heatmap of parameter likelihoods showing the most likely r and ρ values with prior knowledge of division rate, λ, as inferred from histone dilution experiments. a) the analytical engine b) Gillespie simulations with MLE and 100,000 sample size, c) Gillespie simulations with cumulative KS distance and 100,000 sample size, d) SMC-ABC running for 10 generations. Heatmap plots Kernel Density Estimate of final population of parameter sets.

(a)

(b)

(c)

(d)
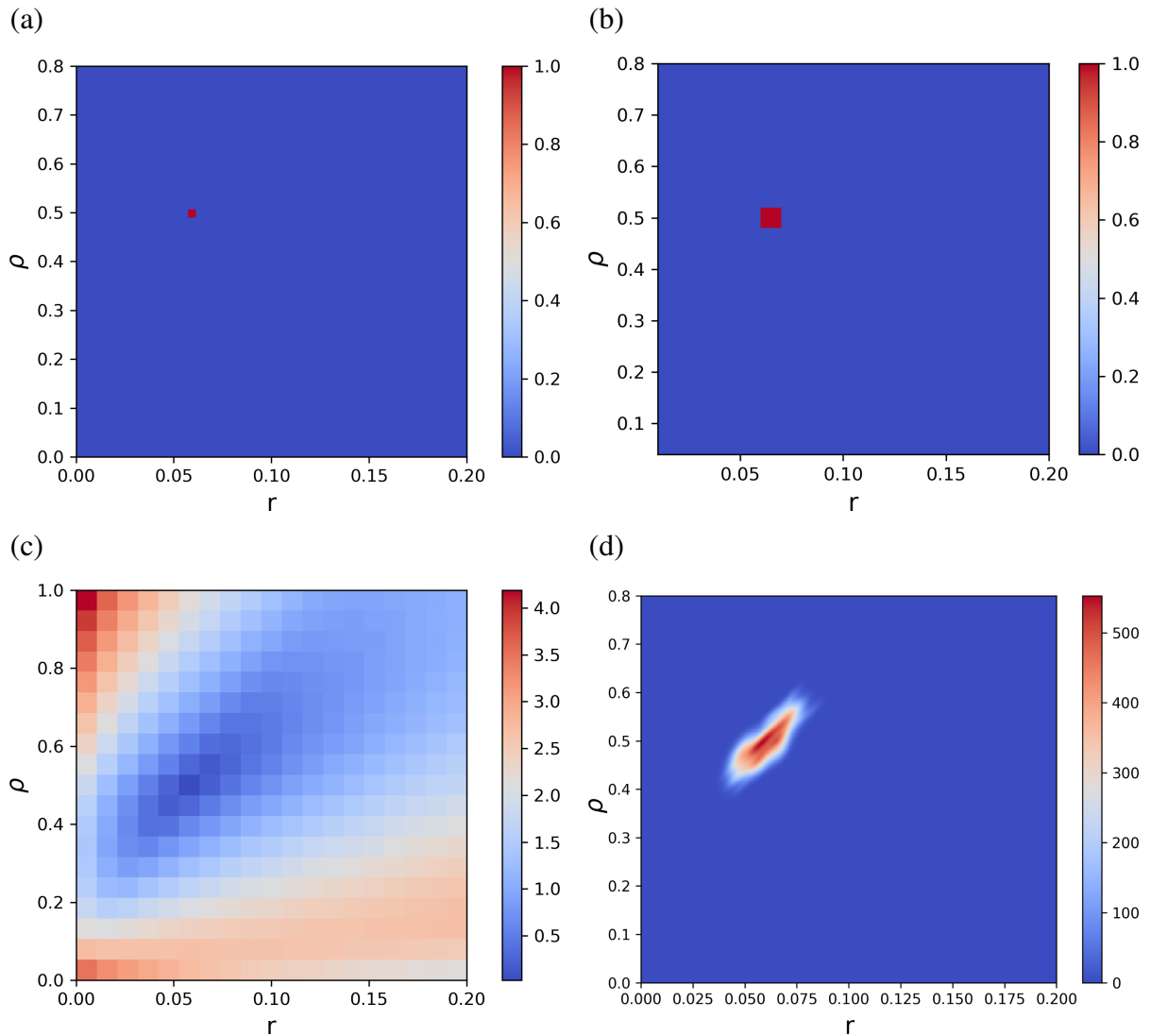
Fig. 3.15 Analytical solution and simulation-based methods attempting to infer SP parameters when applied to synthetic dataset. The synthetic dataset was generated with parameters $\lambda = 2.9$/week, $r = 0.06$, $\rho = 0.5$ and assuming exponentially distributed cell cycle. Heatmap of parameter likelihoods showing the most likely r and ρ values with prior knowledge of division rate, λ, as inferred from a) the analytical engine b) Gillespie simulations with MLE and 100,000 sample size, c) Gillespie simulations with cumulative KS distance and 100,000 sample size, d) SMC-ABC running for 10 generations. Heatmap plots Kernel Density Estimate of final population of parameter sets.

(a)

(b)

(c)
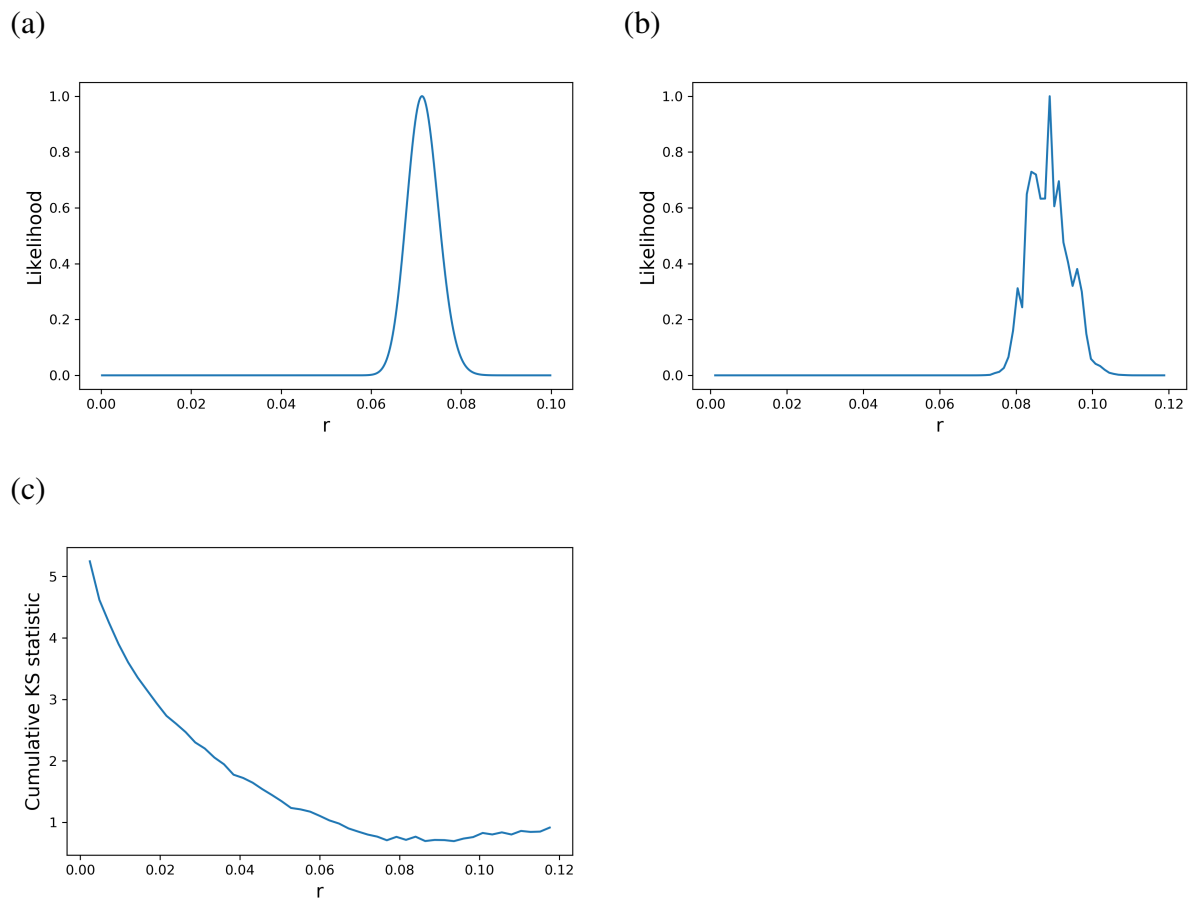
Fig. 3.16 Restricted search assuming a relationship of r/ρ obtained from the linear slope of the average clone size over time. a) Analytical engine gives a smooth distribution of parameter likelihoods with $r = (0.067, 0.0756)$ 95% CI, b) MLE based Gillespie simulations produce a noisier distribution even with increase sample size (1,000,000) with $r = (0.0816, 0.0948)$ 95% CI, c) Gillespie simulations with cumulative KS distance and 10,000 sample size.

3.5.1 Effects of cell-cycle and biological variation on parameter inference

As already stated, both the analytical engine and the simulation based approaches examined so far assumed an exponential distributed cell cycle of stem cell population. However, this assumption does not take into account the waiting time between consecutive cell divisions observed in real tissues. This is also supported by short term data from histone labelling and live imaging which suggest that cell cycle does not follow an exponential distribution. Moreover, the observation of large sized clones in Gillespie simulations I performed, which are not frequently observed in lineage tracing further indicates that real data follow alternative cell cycle distributions. Furthermore, the original synthetic datasets do not take into account

biological variation between mice (Sada et al., 2016, Piedrafita et al., 2020), and so do not allow testing whether error estimates are accurate. To test these properties explicitly both features were included in synthetic datasets and rerun. I initially generated synthetic data with a more realistic underlying cell cycle distribution (Gamma distribution) (Piedrafita et al., 2020):

$$P(t_{cc}) \sim Gam(\kappa, \theta), \tag{3.2}$$

where κ and θ represent the shape and scale parameters of a Gamma distribution.

Application of the analytical method to Gamma distributed synthetic data revealed an issue in inferring the expected parameter values. As shown in (Figure 3.17), although the method seems to accurately infer the expected r and ρ values in the synthetic dataset with exponentially distributed cell cycle, the estimated parameter values were shifted when Gamma distributed cell cycle was considered. This highlights an important limitation of the analytical solution, as it appears to analyze accurately only datasets with unrealistic cell cycle distributions.
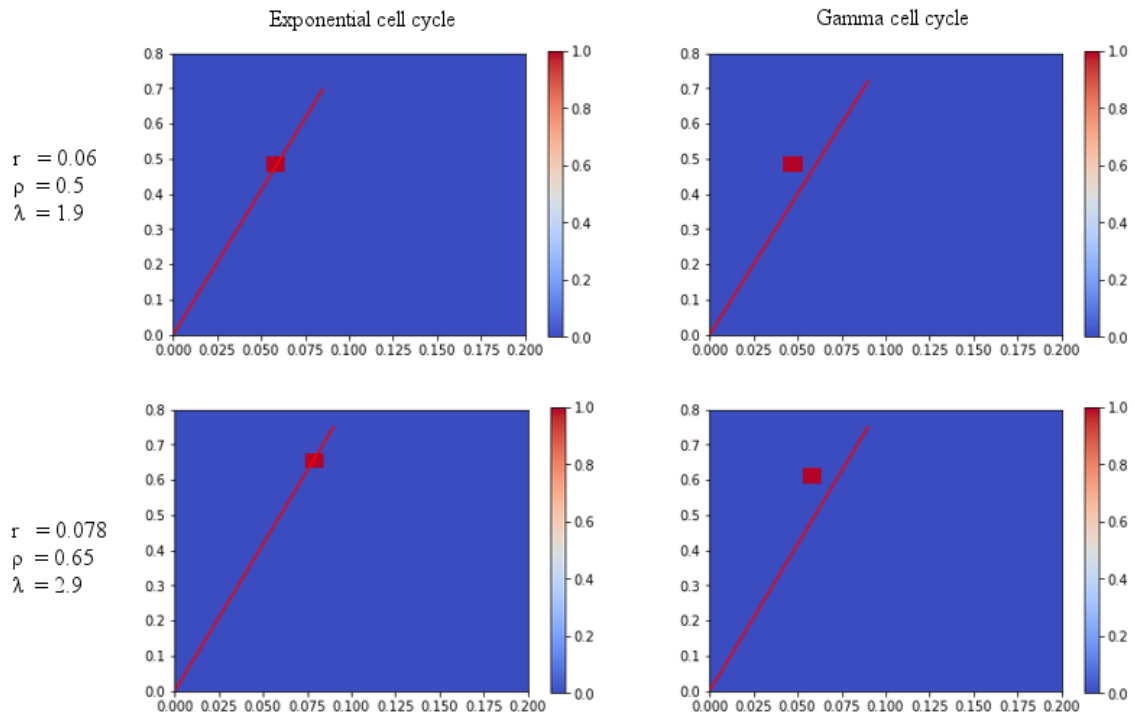
Fig. 3.17 Parameter inference outputs of the analytical solution on synthetic data following either an exponential or Gamma distribution of average cell cycle times . Inference of r and ρ parameters in datasets with Gamma distributed cell cycle times appears to be less accurate. Parameters of Gamma distribution: $\kappa = 8$, $\theta = 0.3875$ (top), $\kappa = 8$, $\theta = 0.2375$ (bottom), from Piedrafita et al. (2020). Red line corresponds to r/ρ ratio obtained from the linear slope of the average clone size over time.

To test whether the inference process is also sensitive to biological variability, I generated synthetic data with r, ρ and λ values drawn from a normal distribution in order to introduce some noise, similar to the potential noise introduced in real datasets by biological variation. Synthetic datasets where realistic cell cycle distribution times and biological variation were modelled revealed that likelihoods calculated from the analytical engine were overly precise with the narrow confidence intervals that appear to underestimate the uncertainty on the maximum likelihood estimate (Figure 3.18a). Whilst the introduction of more complex cell cycle times undermines the maximum likelihood parameters, the underestimation of errors by maximum likelihood methods has been reported as a result of poor sampling (Jain and Wang, 2008, Sugasawa and Noma, 2019). Using a simulation based maximum likelihood approach improves the parameter estimation but error estimation and computational efficiency issues remain. An SMC-ABC approach was able to overcome this issue. This method offered an efficient and more accurate input parameter inference, producing a smooth distribution with more realistic confidence intervals (Figure 3.18b). Together, these findings demonstrate that the SMC-ABC inference technique can be considered as the most appropriate method for analyzing lineage tracing datasets.

(a)   (b)
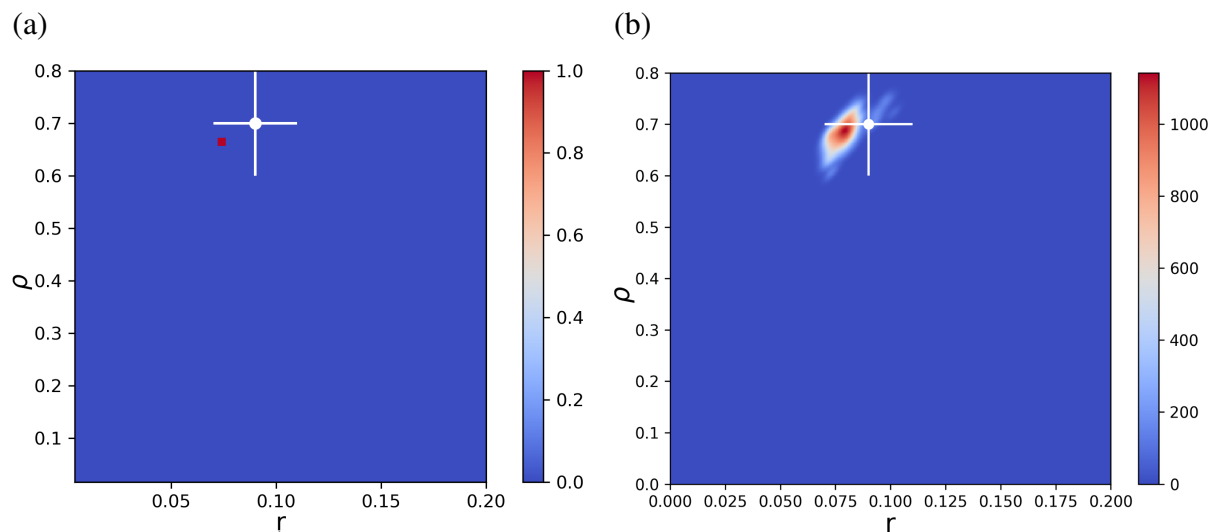
Fig. 3.18 SMC-ABC approach efficiently and appropriately parameterises the single progenitor model from lineage tracing data. a) Synthetic datasets where realistic cell cycle distribution times and biological variation are modelled show that likelihoods calculated from the analytical engine are overly precise with too narrow confidence intervals and thus inaccurate. Input parameters and error estimates indicated by cross and error bars (2XSD), whilst the heatmap plots likelihood distribution calculated using the analytical solution. b) An SMC-ABC approach gives a smooth distribution and reasonable confidence intervals. Input parameters and error estimates indicated by cross and error bars, whilst the heatmap plots Kernel Density Estimate of final population of parameter sets.

3.6   Re-analysis of oesophageal and epidermal datasets with SMC-ABC

Having identified the best approach for analysing clonal datasets, I revisited the data from oesophagus (Section 3.2.3, page 44) and epidermis (Section 3.2.4, page 46) and performed parameter inference using SMC-ABC. The analysis was performed considering Gamma distributed cell cycle times. As demonstrated in Figure 3.19 the method estimated the r and ρ parameters for each dataset, with the predictions being in line with those initially proposed by the analytical solution. However, ABC estimates are not overly precise and have broader confidence intervals. The proposed parameter values are also in broad agreement with previously published ones for the same tissues (Doupé et al., 2012, Murai et al., 2018).
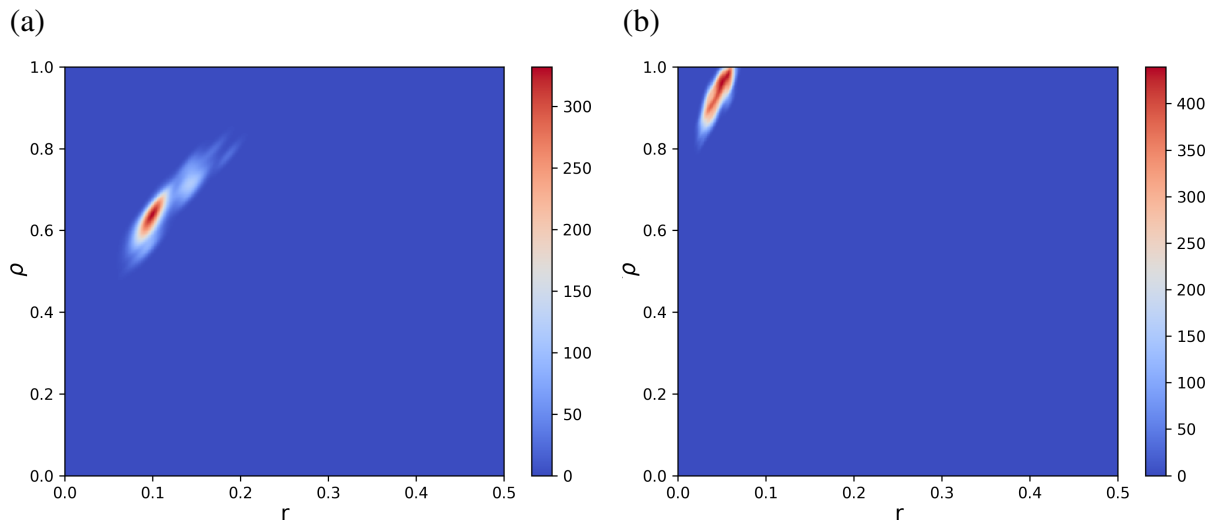
Fig. 3.19 Application of SMC-ABC on oesophageal and back epidermis lineage tracing data. SMC-ABC parameter inference output for a) oesophageal data b) Epidermal data. SMC-ABC estimates parameters with more realistic confidence intervals, in contrast to the narrow ones produced by the analytical solution. The inferred r and ρ appear consistent with previously proposed values describing datasets of the same tissues (Doupé et al., 2012, Murai et al., 2018).

3.7 Discussion

To summarize, in order to investigate how mechanisms that ensure the proper function of normal squamous epithelial tissues are broken during mutagenesis and early development of cancer, one has to primarily focus on how these tissues are maintained. For this purpose, it is important to retrieve quantitative information of the epithelial stem cells dynamics in homeostasis. Analysis of stem cell population distributions at different time points from lineage tracing experiments in oesophageal and epidermal tissues gave rise to a single progenitor model of tissue maintenance. This mathematical model can be described by a set of parameters, a division rate (λ), a stratification rate (Γ) and the probability of symmetric division (r).

An analytical approach combined with maximum likelihood calculations has been proposed for estimating these parameters and fits given clonal data as input. Using this existing solution, short and long term oesophageal and epidermal experimental data were tested. Discrepancies in parameter estimation revealed serious concerns about this method and made its validation on computationally generated data highly significant. Quantitative analysis of the synthetic data allowed me to further examine the issues in the method, fix some of them and discover the major problem of collision in estimated parameter probability distributions. This would imply that we are not able to do better fitting to the distributions than looking at their average properties. Thus, in order to estimate all three parameters, it is necessary to calculate at least two of them from alternative sources.

Considering the above issues of the current computational approach in parameter estimation I sought to find a more accurate and efficient way to quantify stem cell clone dynamics. To this end a series of different strategies were explored on both experimental and synthetic datasets. Re-implementation of the analytical solution in Python led to the elimination of estimating same probabilities for different parameters. The new analytical engine was able to infer parameters with narrow confidence intervals. Model parameterisation and accurate representation of uncertainty however remained problematic, when realistic cell cycle distributions and biological variation between animals were explicitly included. In particular the parameters proposed were incorrect and the true values outside the confidence intervals.

Simulation maximum likelihood approaches had the advantage of taking account of more realistic cell cycle distribution times but require intractable numbers of simulations to estimate the most likely parameter combination accurately. A substantially faster and computationally cheaper ABC approach was able to account of cell cycle times and identify the input parameters more accurately with reasonable error estimates.

Taken together, the major output from the above analysis was to identify the most appropriate approach for analysing lineage tracing data on squamous epithelia. This method will allow for accurate and efficient analysis of newly collected datasets. Most importantly, reliable parameterisation of such systems would give access to valuable quantitative information on how these tissues are maintained and enable investigating what causes homeostatic disruption during mutagenesis and cancer.

Chapter 4

Spatial single progenitor model continues to describe epithelial homeostasis

Abstract

Homeostasis in stratified squamous epithelia is dictated by the coordinated behaviour and neutral competition of stem cells in the context of a tightly packed tissue. The SP model successfully reproduces epithelial stem cell dynamics in the zero dimension. In this chapter, I sought to confirm that such a model continues to match experimental data when spatial competition is explicitly included. To address this question, I developed a spatial SP model using Cellular Automata (CA).

I found that the spatial SP model successfully recapitulated the hallmarks of neutral growth. However, the cell autonomous fate decision process supported by the SP model, introduced local, transient crowding and extinction events, indicating the need for cell communication feedbacks to correct the local imbalance.

Nevertheless, coupling division and stratification events as an attempt to introduce some form of spatial regulation did not offer a substantial improvement to the simulated dynamics of the system. On the contrary, in some cases, fate coupling failed to reproduce the key homeostatic properties.

Together, these findings suggest that tissue organisation does not alter the neutral SP model as the growth patterns of the spatial model accurately reproduced the experimental observations across different available experimental datasets.

4.1 Introduction

The study of cell fate determination and the balance between cell production and loss underlying tissue homeostasis, provides means to understand both tissue maintenance and carcinogenesis in stratified squamous epithelia. The single progenitor (SP) model has proven particularly useful as it offers parameterisation and a quantitative description of epithelial tissue maintenance.

However, such a model permits the study of stem cells as populations and completely disregards the spatial characteristics of the tissue. Spatial competition plays an important role in a wide variety of different scenarios, altering growth curves in ways that reflect both tissue structure and experimental methods (Noble et al., 2019, Hall et al., 2019, Chkhaidze et al., 2019). Tissue organisation restricts clone growth, potentially altering growth patterns and undermining the validity of the model. In the single progenitor model specifically we might expect that average clone size grows initially linearly, but becomes sublinear as clones grow large and only can expand through competition with neighbouring clones at the periphery. Thus, to investigate whether the inclusion of spatial constraints alters growth dynamics, the SP model was extended to account for the spatial patterning of cells within the basal layer.

Several spatial modelling approaches have been used to investigate processes such as healthy tissue growth and maintenance, as well as tumour growth and wound healing in multi-cellular systems. In such cell-based or individual-based models, cells are represented as discrete entities with assigned properties, which respond to local interaction rules. Depending on the information required to be captured, different model classes can be used, with different levels of complexity. Two main categories of cell-based model types exist: on and off lattice, based on whether cells are restricted to a fixed grid or they move freely in response to applied forces (for a more detailed description, see Section 1.9.2).

Two well known classes of on-lattice cell-based models are cellular automata (CA) and the cellular Potts (CP) model, where a biological cell may occupy a single or several lattice sites respectively. In CA, the system evolves according to a set of local rules, which specify how each site is updated (e.g. division, migration, death). The low complexity of this model class, allows for performing multiple simulations of large cell population sizes efficiently. In contrast to CA, the CP modelling approach provides a more realistic representation of cell shapes and is capable of studying mechanical properties of the system, such as cell adhesion and membrane tension (for a more detailed description, see Section 1.9.2).

A more detailed exploration of mechanical processes is achieved with the off-lattice or lattice-free models. This model class, where cells are usually represented as deformable

spheres which change shape following interactions, or as polygons whose vertices move according to a balance of forces, achieves a higher resolution in cell shape and complex cell packings (for a more detailed description, see Section 1.9.2). Lattice-free simulations offer a natural approach for considering crowding without further modification. However, this approach relies on additional parameters which cannot be retrieved by the available lineage tracing data. Therefore, an on-lattice model would be a more appropriate modelling approach for simulating the current system.

The available lineage tracing data provide information regarding clone counts and cell positions within the tissue without however including physical properties in detail. Considering this, a simpler model, which keeps track of each cell's decisions without explicitly accounting for mechanical interactions is more appropriate for describing the current experimental information. Therefore, a stochastic CA was used to implement the SP model in the two-dimensional space and explore the collective behaviour of cells in the tissue. The appropriateness of CA in modelling lineage tracing experiments is also supported by a recent study (Roy et al., 2014).

4.2   Spatial Single Progenitor model

Experimental images of mouse epidermis and oesophageal epithelium have shown that cells are organized in a hexagonal-like arrangement (Figure 4.1), leading me to select a hexagonal lattice as a reasonable approximation of an adult mouse homeostatic epithelial basal layer. A two-dimensional, hexagonal lattice was used to model the basal layer of the epithelium (Figure 4.2a). Each site of the grid may be occupied either by one of the two cell types described in the SP model (Section 1.5, page 8) proliferating cells (A) and post-mitotic cells (B), or it may remain vacant as a result of a stratification event. Also, a lattice site may be occupied by two cells, indicating a crowding region. A division event can lead to three potential outcomes: two proliferating cells, two differentiating cells or one daughter of each cell type.

The neighbourhood in the SP CA model is defined by the six adjacent places (Figure 4.2b). Each newly born cell can be placed on any vacant space in its parent's six-cell neighbourhood. However, lineage tracing experiments on the murine oesophageal epithelium reveal that cell divisions are aligned within the longitudinal axis of the organ (Doupé, 2009). This preferential direction of divisions might be possibly explained by the continuous longitudinal tension exerted on the oesophageal epithelium, by means of its attachments to the oral cavity and the stomach (Doupé, 2009). To explicitly include the above mentioned reported feature

in the model, a division directionality bias was introduced for simulating the oesophageal epithelium dynamics. The bias was introduced by defining as neighbourhood the four adjacent places instead of six, thus forcing division events to take place along one axis of the grid (Figure 4.2c).
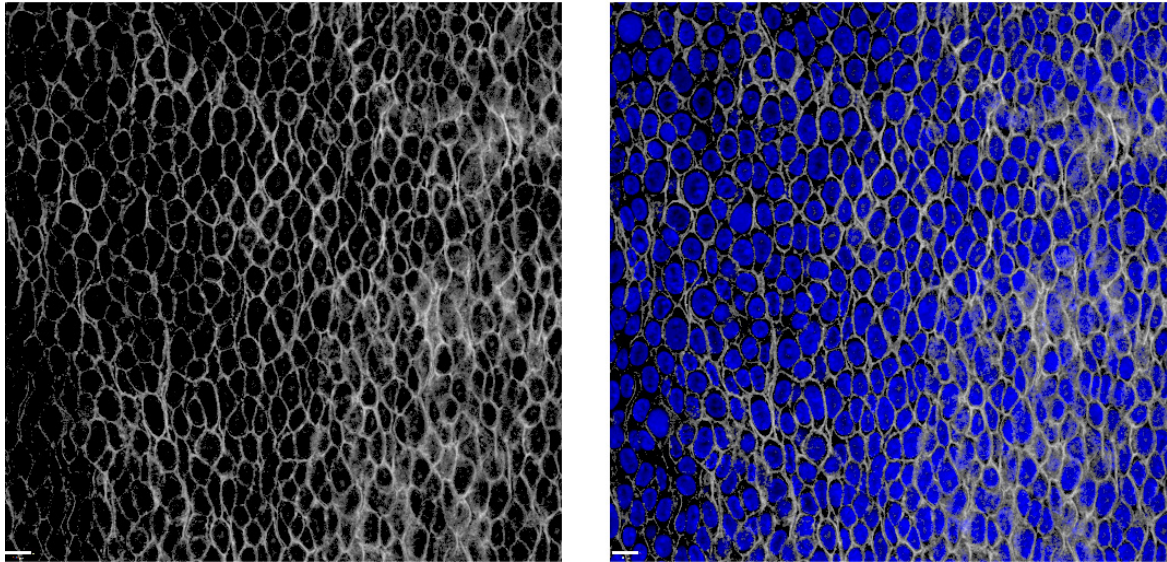
Fig. 4.1 Confocal images of mouse OE basal epithelium showing a hexagonal-like cell organization. Left: Grey is Wheat Germ Agglutinin (WGA) membrane staining. Right: WGA (grey) showing cell's membranes and DAPI (blue) showing cells' nuclei. Scale bars, 9 μm. Image provided by Phil Jones lab, in the Sanger Institute.
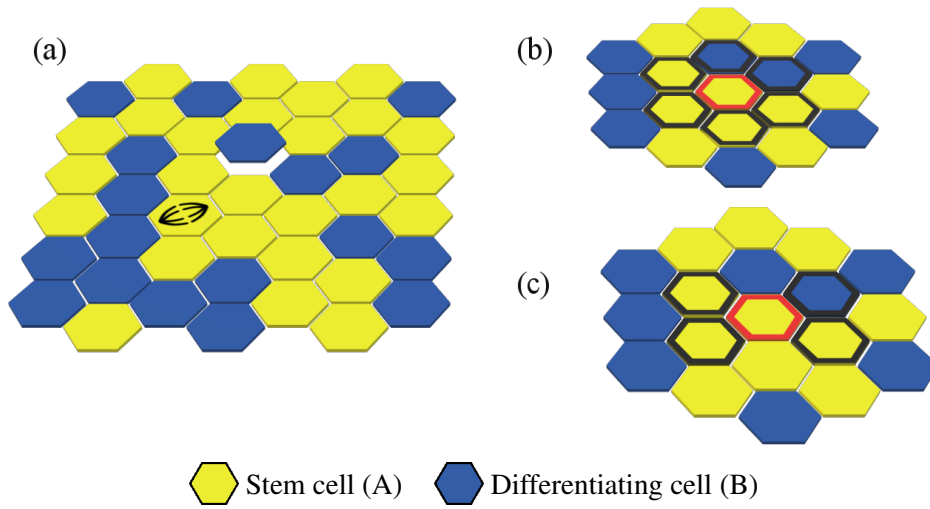
Fig. 4.2 Schematic representation of the spatial SP model. a) illustration of the two-dimensional hexagonal lattice representing the epithelial basal layer. Proliferating cells (shown in yellow) are able to divide, whilst differentiating cells (shown in blue) exit the basal layer and are removed from the simulation, b,c) neighbourhood types: stem cells can divide and the new daughters can be placed on any of the six adjacent lattice sites (b) or on any of the four adjacent sites to simulate the preferential direction of divisions along the longitudinal axis in the murine oesophageal epithelium (c).

The spatial single progenitor model was initially developed without accounting for any kind of spatial feedbacks that could influence cell fate decisions, similar to the non-spatial model. Therefore, division and stratification events were considered as two independent processes determined solely by r, λ and Γ parameters. This cell-autonomous approach could lead to cases where a cell division event occurs at a region with no available neighbouring vacant space. In such case, the two daughter cells were placed on the same grid space, indicating an increased cell density area. Analogously, cases where an empty space generated by a recently stratified B cell is not rapidly replaced by a nearby newly born cell might be observed, representing a low cell density area. Considering the above, each lattice site could have one of the following seven potential states: A, B, D_{AA}, D_{AB}, D_{BA}, D_{BB} and "empty" ($\varnothing$), where D_{AA}, D_{AB}, D_{BA}, D_{BB} correspond to double occupancies. Thus, the SP model (Eq 1.1, page 9) was extended to explicitly include space as follows:

$$A \varnothing \xrightarrow{\lambda} \begin{cases} AA & r \\ AB & \frac{1}{2} - r \\ BA & \frac{1}{2} - r \\ BB & r \end{cases}$$

$$A X \xrightarrow{\lambda} \begin{cases} D_{AA} & r \\ D_{AB} & \frac{1}{2} - r \\ D_{BA} & \frac{1}{2} - r \\ D_{BB} & r \end{cases} \tag{4.1}$$

$$B \xrightarrow{\Gamma} \varnothing,$$

where $A \varnothing$ denotes a type A cell neighbouring a vacant lattice site and AX denotes a type A cell neighbouring either a type A or type B cell, thus indicating that there is no neighbouring empty space. D_{AA}, D_{AB}, D_{BA}, D_{BB} correspond to double occupancies. The choice to include double cell occupancies in the model was based on *in vitro* culture observations (Phil Jones, personal communication). Greater than double occupancies were not observed in *in vitro* cultures and were not included in the model.

Figure 4.3 represents a schematic description of the spatial rules of the SP CA model.
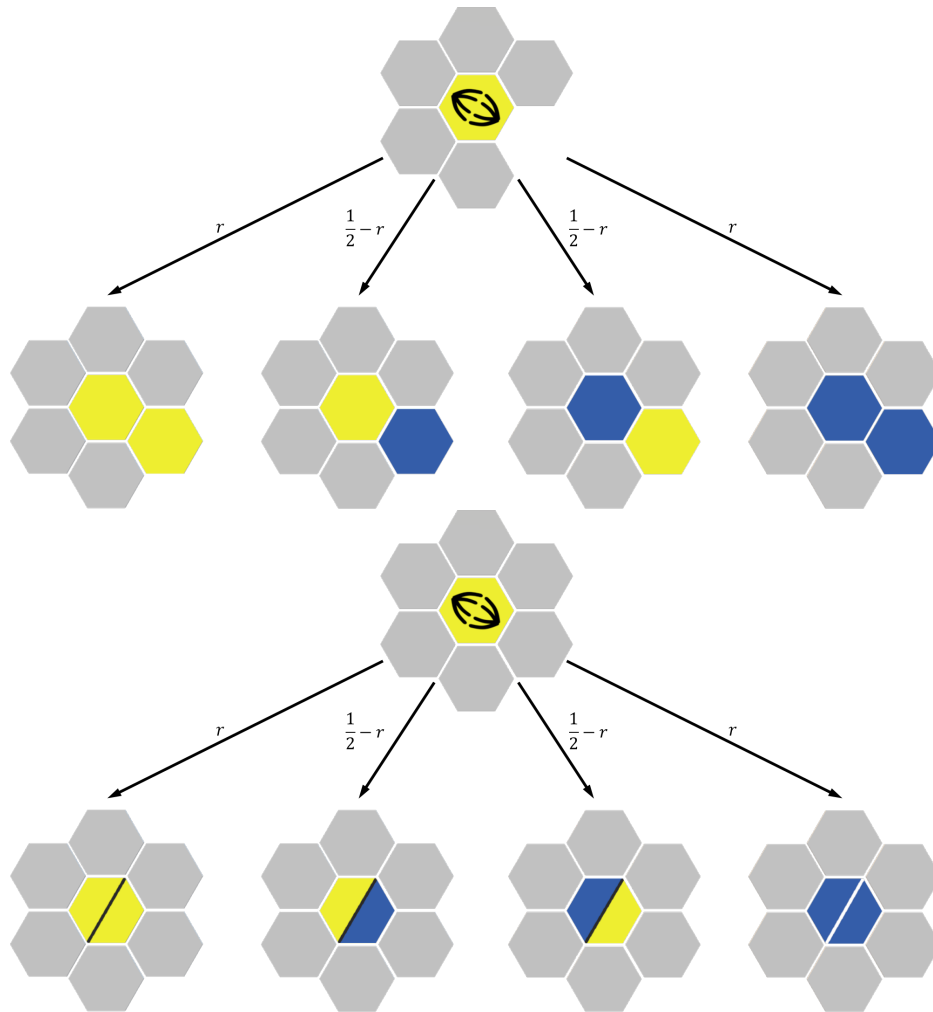
Fig. 4.3 Rules of the spatial SP model. Illustration of the spatial SP model. Proliferating cells in yellow and differentiating cells in blue. Proliferating cells undergo a division type which is decided by the probability of symmetric divisions,r. A proliferating cell which is about to divide checks its immediate neighbourhood for available space. If a vacant site exists (top), one daughter cell occupies the mother cell's space and the second the neighbouring empty space. If there is no empty space in the immediate neighbour (bottom), the two daughters occupy the mother cell's space, thus creating a double cell occupancy. Double state cells are released once a neighbouring lattice site becomes available.

The CA model was developed in NetLogo, a widely used agent-based modelling environment (Wilensky, 1999). To investigate how spatial constraints affect epithelial cell clone stochastic evolution, I used a Markovian stochastic simulation algorithm, where the basal layer was simulated as an asynchronous CA. The algorithm included the following steps:

1. Start by defining a grid of *NxN* sites with A and B cells randomly seeded. The fraction of A cells is calculated from ρ parameter.

2. For every cell on each lattice site, draw a random number from an exponential distribution with mean $1/\lambda$ or $1/\Gamma$ to assign time of next event (division or stratification) for A and B cells respectively.

3. Select cell with the smallest next event time assigned. Current time is updated to the smallest next event time.

4. If an A cell is selected, use a random number from a uniform distribution $U \in (0,1)$ to choose the division type to occur by comparing U to division probabilities. Assign the division type as a next event for the selected cell. If a B cell is selected, assign stratification as a next event for the selected cell.

5. If the next event is division, all neighbouring places are checked for empty space. In the case of an existing neighbouring space, one new born cell will replace the mother cell and the other will occupy the empty neighbouring space. If there is no empty neighbouring space available then both will remain at the mother cell's space (creating a "double state" cell), until a neighbouring space is released. This would cause delays in division and stratification events as "double state" cells are trapped and thus not able to divide or stratify at their assigned times. If stratification is the next event, B cell stratifies, leaving an empty space, which allows potential neighbouring "double state" daughters to be released.

6. Repeat steps 3-6 until there are no A or B cells left or time threshold is reached.

The following flowchart describes the model steps in detail:
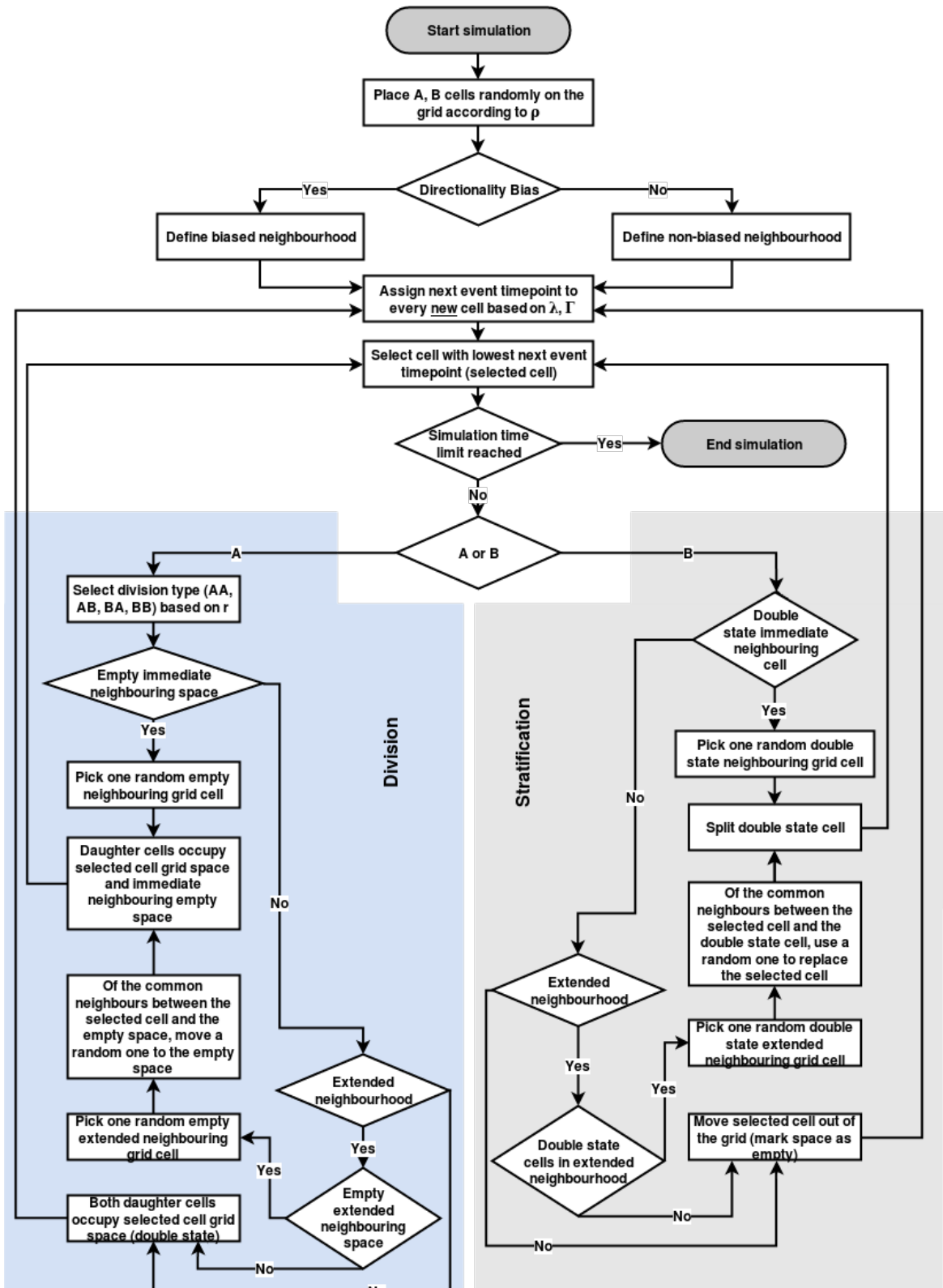
Fig. 4.4 Spatial Single Progenitor model. Detailed flowchart describing the steps of the spatial SP model.

4.3   SP spatial simulations of WT oesophageal epithelium

Computational simulations of the CA model were performed to explore the dynamics of the spatial system. The mouse oesophageal epithelium is considered a well-studied system, mainly due to its simple, homogeneous architecture and rapid regeneration. We can assume that the published r, λ and ρ parameters inferred from quantitative analysis of murine oesophageal lineage tracing data are reliable. For this reason, the spatial simulations were set up to model the dynamics of the epithelium, using the input parameters provided by the literature. Basal epithelial cells were simulated on a lattice originally containing 10,000 cells ($L = 100\ X\ 100$), corresponding to roughly 1% of adult mouse oesophagus, where periodic boundary conditions were applied. Each simulation was repeated 100 times.

To assess how many simulation repeats are required in order to achieve output convergence, I performed several simulation repetition sets (N=10, 30, 50, 100, 200) and tested how reliably each of them was able to predict a known property of the system. As a known feature, I used the slope of the average clone size over time, which should equal $r\lambda/\rho$ (as described formally in Section 1.5). As shown in Figure 4.5a, all tested repeat sets consistently predicted a slope value equal or similar to the expected one.

To adopt a more systematic approach in identifying the required number of simulation repeats, I performed a Jackknife re-sampling analysis (Efron and Stein, 1981), calculating the mean clone size statistical value for an increasing number of repetitions. Jackknife is a technique that performs re-sampling (data reuse for generating new, hypothetical subsamples) to evaluate the bias and variance of a parameter estimate as well as calculate the standard error of a statistic. The new samples are generated by systematically excluding each observation from the original data and a parameter estimate is then calculated for each sub-sample. As expected, as the number of repetitions rises, the standard error of mean clone size decreases. However, this decrease is sharper from 10 to 100 repetitions and it slows down when moving towards 200 repetitions (Figure 4.5b). In addition, more than 100 repetitions result in an increased demand in computational resources with a negligible improvement in standard error. Considering all the above, 100 repetitions per simulation were selected.

Given that the mouse lineage tracing experimental data available typically provide information for up to 18 months, reflecting mouse lifespan, the duration of the simulations was set to 80 weeks.

(a)   (b)
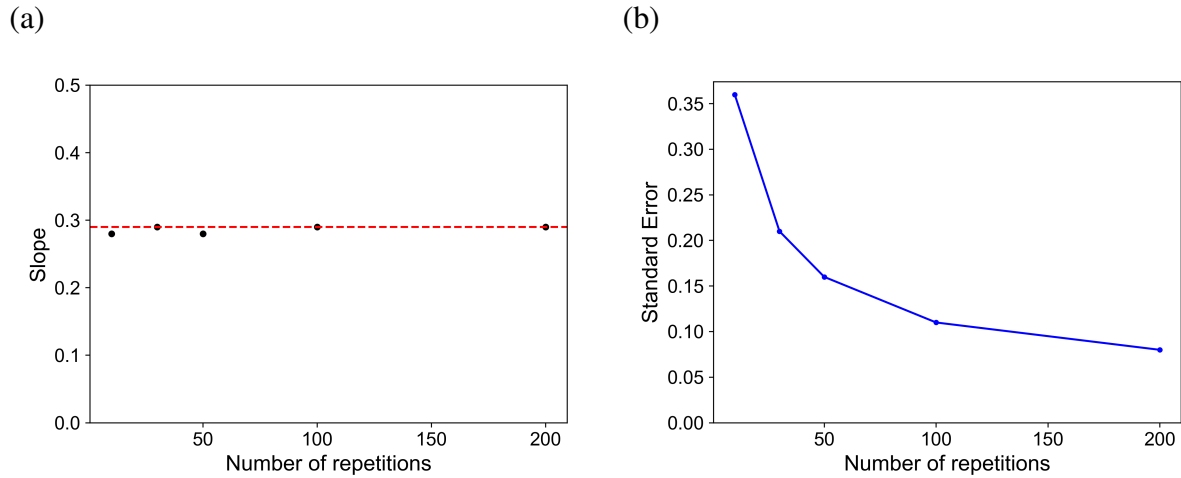
Fig. 4.5 Convergence analysis (a) and Jackknife re-sampling (b) for mean clone size parameter. The average clone size was calculated from spatial SP simulations, repeated 10, 30, 50, 100 and 200 times on a 100x100 lattice with parameters $r = 0.1$, $\rho = 0.65$, $\lambda = 1.9$/week, from Doupé et al. (2012). a) All tested repeat sets consistently predicted a slope value (black points) equal or similar to the expected one (red dotted line). b) Standard error of mean clone size for the different simulation repetition sets as estimated by Jackknife re-sampling. An initial sharp drop is observed which slows down as the number of repetitions increases.

4.3.1 Single Progenitor hallmarks on cell populations are successfully recapitulated

The simulation outputs revealed that taking into consideration the spatial constraints imposed by the lattice do not cause the model to deviate from the experimental data taken from mouse oesophagus. The spatial model is able to reproduce the characteristic features of the stochastic birth-death process (the hallmarks of a single population of stem cells), as described in Section 1.5. Homeostasis (a constant number of proliferating cells) is maintained by a balance between cell production and loss (Figure 4.6a). Over time, due to stochastic divisions resulting in a pair of differentiated daughter cells, clones are lost following a simple relationship determined by r and λ. which is counterbalanced by a continuous increase in their size. To maintain a constant population, the average size of the persisting clones rises linearly, with slope $\tau = r\lambda/\rho$. The simulated tissue demonstrates both of these properties (Figure 4.6b,c) with growth dictated as expected by input model parameters. Finally, whilst the clone size distribution becomes broader over time, the shape remains constant once scaled by the average clone size (Figure 4.6d,e). Taken together, these findings demonstrate that the average behaviour and the clone size distributions of the epithelial basal cells are consistent with the stochastic SP model of epithelial homeostasis, implemented on the two-dimensional space.

(a)

(b)

(c)

(d)

(e)
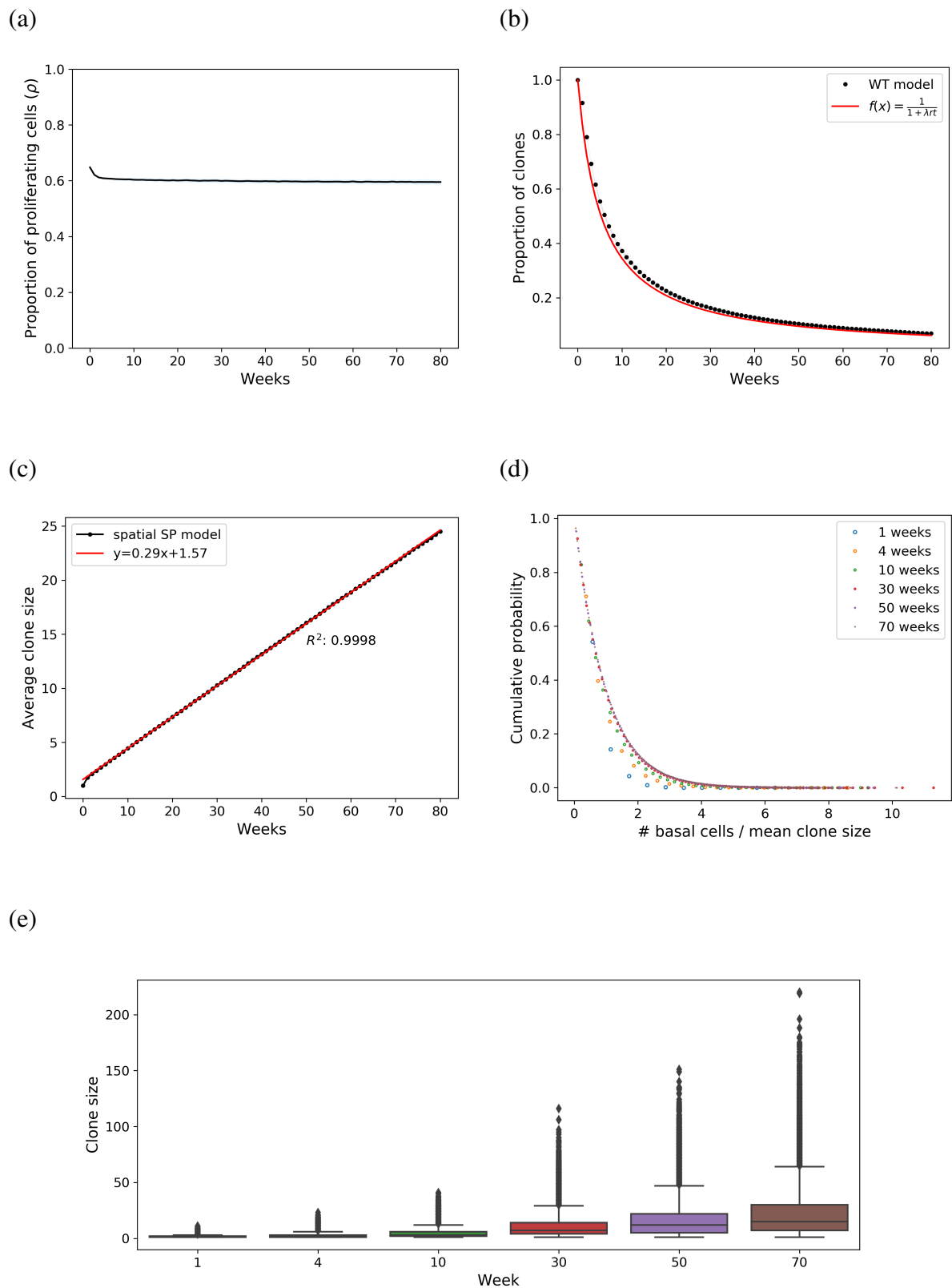
Fig. 4.6 Quantitative analysis of spatial SP simulations for the oesophageal dataset. Parameters used: $r = 0.1$, $\rho = 0.65$, $\lambda = 1.9$/week, from Doupé et al. (2012). Spatial restrictions do not alter the characteristic features on cell population dynamics, described in Section 1.5. a) stem cell population remains largely constant, b) the number of clones decreases over time following a simple relationship determined by r and λ, c) the average clone size increases over time, with slope $\tau = r\lambda/\rho$, d) clone size distribution scales over time, e) the clone size distribution adopts a broader shape over time. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

4.3.2 Spatial impact on tissue density

Murine oesophageal tissue presents a remarkably simple architecture, with no complex structural units and a nearly uniform cell density (Doupé et al., 2012, Alcolea et al., 2014). To investigate whether the spatial model is able to reproduce the experimentally observed uniform basal layer density, I examined the spatial distribution of cells in the grid. The spatial component introduces some interesting emergent properties. Several random crowding and extinction events are noticed, demonstrated by the formation of double cell occupancies and empty areas respectively (Figure 4.7). These patterns emerge as small-scale, local events on different parts of the lattice as time progresses.
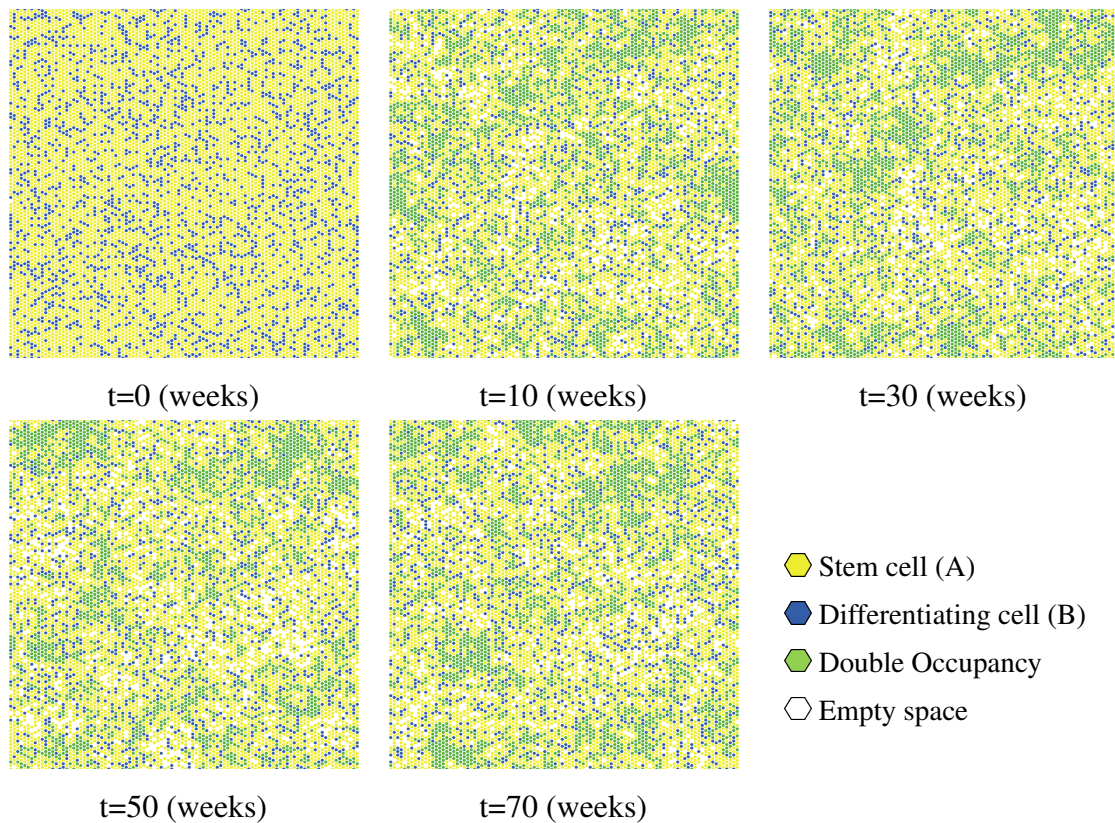
Fig. 4.7 **Spatial SP simulation time lapse.** Typical simulation time lapse of the spatial single progenitor model, simulating cell growth dynamics in mouse oesophageal basal layer at different time points (0, 10, 30, 50 and 70 weeks). Cells are placed on a 2D 100x100 hexagonal lattice. The inclusion of space introduces cell density heterogeneity. Areas of empty patches (shown in white) and double state cells (shown in green) are formed, as local transient phenomena. Parameters used: $r = 0.1$, $\rho = 0.65$, $\lambda = 1.9$/week, from Doupé et al. (2012).

(a)
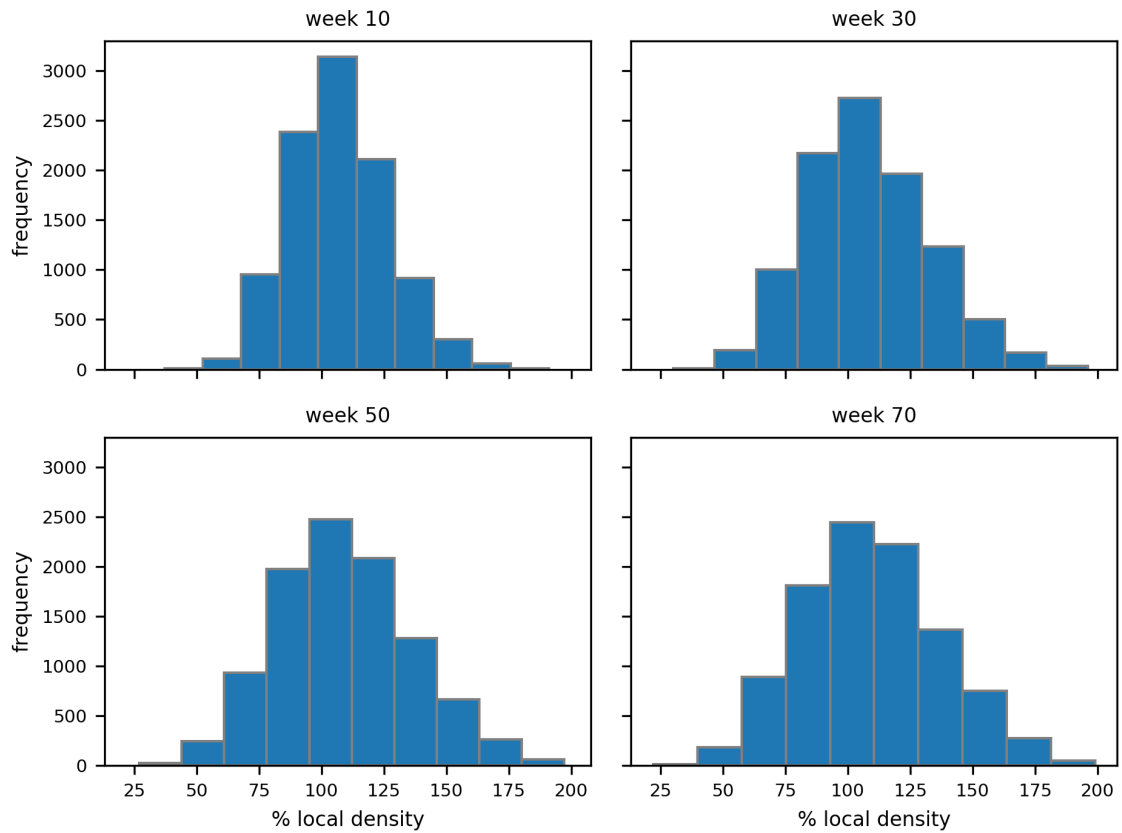
(b)
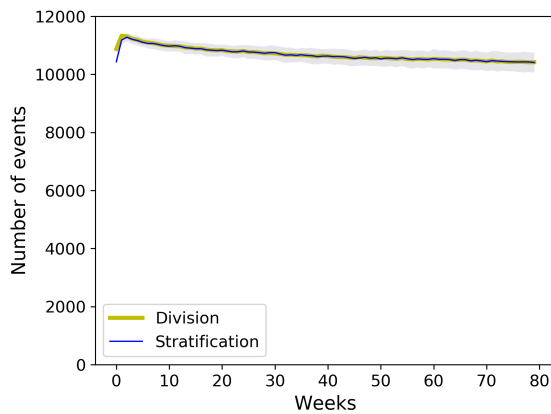
(c)
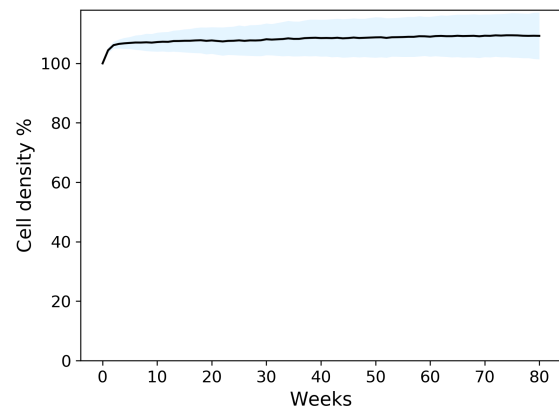
Fig. 4.8 Effects of cell density fluctuations on spatial SP simulations of mouse oesophagus. Parameters used: $r = 0.1$, $\rho = 0.65$, $\lambda = 1.9$/week, from Doupé et al. (2012). a) Local cell density distribution across equally sized lattice segments adopts a broader shape over time, indicating an increase in cell density heterogeneity within the simulated tissue. Grid was split to 100 segments of 10x10 size, b) the emergence of high density regions leads to a slight decrease in division and stratification events over time, c) tissue overall basal cell density remains largely constant over time. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

As a consequence, cell density fluctuations occur, rendering local cell density distribution broader at later time points (Figure 4.8a), generating either jammed regions of trapped cells which are not able to proliferate, or gaps. These local density fluctuations have an effect on proliferation and stratification rates which follow a slightly decreasing trend (Figure 4.8b). However, the spatial simulations are able to produce an overall normal adult oesophageal basal layer, as demonstrated by the mostly constant basal cell density (Figure 4.8c). Taken together, the local, transient phenomena in cell density variations, not previously quantified in real tissues, indicate that tissue structure might limit cell growth.

4.3.3 Quantitative comparison to experimental images

In seeking to address the question whether the previously indicated model density fluctuations are realistic, I searched for potential cell density variations in the existing *in vivo* lineage tracing images by comparing their cell density distribution to the simulated ones (Figure 4.8a). A lower variability in the real tissue would result in a less broad distribution. For this purpose, I analysed confocal microscopy images of murine oesophagus wholemounts, provided by Phil Jones in Sanger Institute. The tissue samples were retrieved from a one-year old mouse.
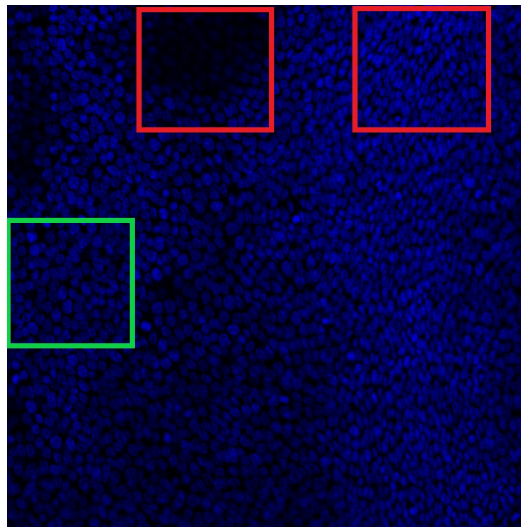
Fig. 4.9 Confocal image of an one-year old mouse oesophageal basal layer. Sample processing for wholemount preparation may lead to increased heterogeneity in the tissue's cell density, which is evident by the formation of crests and valleys. For performing cell density comparison between simulations and tissue images, appropriate regions of confocal images were included. Regions in red squares represent crests and valleys formations and are examples of excluded areas. The region in the green square is an example of an area included in the comparison analysis.
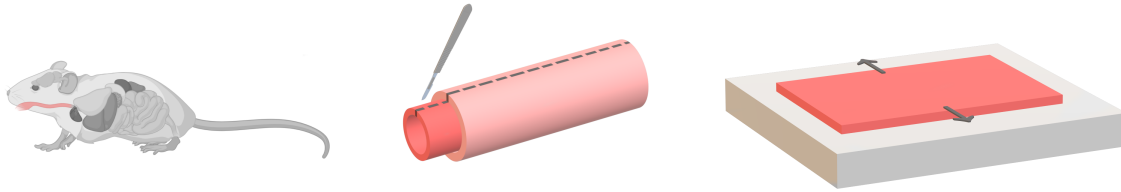
Fig. 4.10 Schematic illustration of oesophagus processing for wholemount preparation. Oesophagus is removed from the animal and cut longitudinally. The epithelium is then peeled away from underlying tissue, opened and stretched for processing.

Despite the fact that the tissue cell density is mostly uniform, its structure may contain formations such as crests and valleys as a result of sample processing for wholemount preparation (Figure 4.10). Due to these morphological properties, the spatial distribution of basal stem cells varies across images (Figure 4.9). This structural heterogeneity is not related to the density variability observed in the computational simulations. In order to avoid misleading conclusions, such regions were not taken into consideration in the comparison of cell density between *in vivo* experiments and computational simulations.

The analysis consisted of three main parts:

1. Segment experimental microscopy images in smaller compartments and exclude problematic compartments (as described above).

   A set of confocal microscopy images (141) of murine oesophageal basal layer were obtained. Each image was segmented to 64 equally sized squares, resulting in 9024 compartments in total. From this set, 400 regions were manually annotated and split in two groups: "problematic" ($n = 200$) and "non-problematic" ($n = 200$). Each group was further split into training and test sets, with a 70:30 ratio respectively, and used as input in a machine learning algorithm. The algorithm was implemented in Python, using ImageAI, a python library with self-contained Deep Learning and Computer Vision capabilities (Moses and Olafenwa, 18). The resulting trained model for image filtering which achieved 84.4% accuracy, was then used to classify the full set of images in the two categories ("problematic", "non-problematic"). Out of the 9024 regions from the initially segmented images, 1076 were predicted as "non-problematic" with $P > 70\%$ and included in the analysis.

2. Computationally count cell numbers in each compartment.

The number of cells per image was counted using Mahotas python library for image processing. This library offers a set of image manipulation functions which have the ability to detect sufficiently individual cells even when they form small aggregates (Figure 4.11). When comparing the automated cell counts against a subset of images with manually counted cell numbers, the algorithm achieved an 8% error.
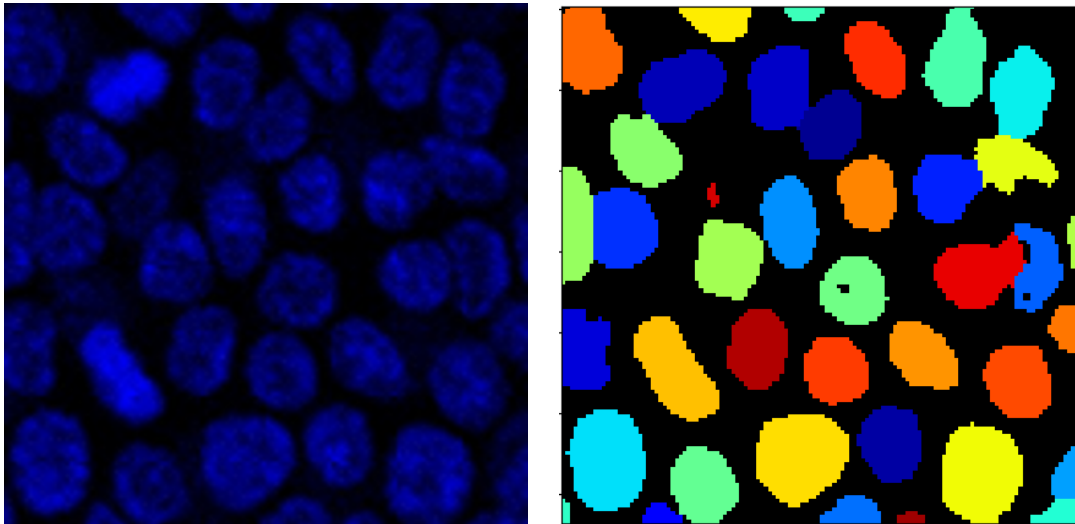
Fig. 4.11 Computational cell counting in segmented confocal microscopy images of one-year old mouse oesophagus. Left: example of an area classified as "non-problematic", which was used as input for automatic cell counting. Right: image processing and cell labelling using Mahotas Python library.

3. Calculate cell density distribution across compartments and compare to computational simulations.

The cell density distribution of the microscopy images and computational simulations were analysed to assess differences in density variance.

As shown in Figure 4.12 the cell density distribution in experimental microscopy images appears less broad, indicating less dispersion in the real tissue. Using a two sample KS test, the two density distributions were shown to be significantly different ($***p < 0.001$), which is further supported by a substantial difference in the coefficient of variation (CV) statistical values ($CV_m = 0.11$, $CV_s = 0.25$, where m denotes microscopy images and s simulations).
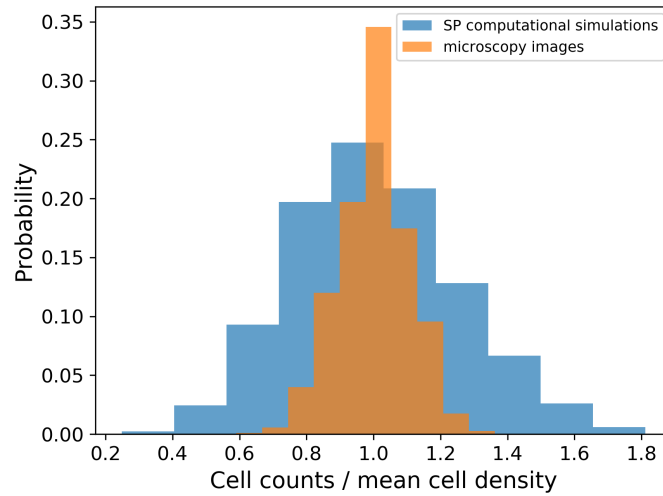
Fig. 4.12 Cell density distribution comparison between experiments and spatial SP simulations. Cell density distribution in confocal microscopy images of murine oesophagus wholemounts and in spatial SP model simulations (t = 52 weeks). To measure local cell density, the lattice was segregated in 100 segments of 10x10 size. Cell counts are normalized for direct comparison. The cell density distribution in microscopy images appears less broad, implying less heterogeneity in the real tissue.

Therefore, the heterogeneity in the spatial distribution of cells, produced by the spatial implementation of the SP model is not observed in the tissue's microscopy images. This would indicate that there might be some kind of spatial regulation mechanism which contributes to the maintenance of a uniform cell density observed experimentally.

4.4   Single Progenitor model variations

4.4.1   Linking division to stratification

The SP model spatial simulations along with cell density analysis of oesophageal microscopy images revealed that a stochastic model of epithelial homeostasis introduces some degree of local cell density fluctuations in the two-dimensional space. Previous studies focused on epithelial tissues provided some evidence of cell behaviour coordination by suggesting that the increased local density introduced by cell division events is compensated by the extrusion of nearby cells (Eisenhoffer et al., 2012, Marinari et al., 2012, Miroshnikova et al., 2018). In light of this, the SP model was modified to consider spatial influences on stem cell stochastic behaviour. The alternative model assumes that division and stratification events are no longer two independent processes, but they are directly coupled. More specifically, the stochastic

division of a stem cell triggers the stratification of a neighbouring post mitotic cell. The division driven model was implemented as follows:

1. Start by defining a grid of *NxN* sites with A and B cells randomly seeded. The fraction of A cells is calculated from ρ parameter.

2. Draw a random number exponentially distributed with mean $1/\lambda$ to assign time of next event (division) for each A cell.

3. Select A cell with the smallest next event time assigned. Current time is updated to the smallest next event time.

4. Use a random number from a uniform distribution $U \in (0,1)$ to choose the division type to occur by comparing U to division probabilities. Assign the division type as a next event for the selected cell.

5. All neighbouring places are checked for a B neighbour. In the case of an existing B neighbouring space, B cell stratifies, one new daughter cell will replace the mother cell and the other will occupy the recently stratified B neighbouring space. If there is no B cell in the neighbourhood, both daughter cells will remain at the mother cell's space (creating a "double state" cell), until a neighbouring B cell stratifies. If a "double state" cell exists in the neighbourhood of a newly produced B cell, the B cell stratifies, allowing the "double state" daughters to split.

6. Repeat steps 3-6 until there are no A cells left or time threshold is reached.

The following flowchart describes the model steps in more detail:
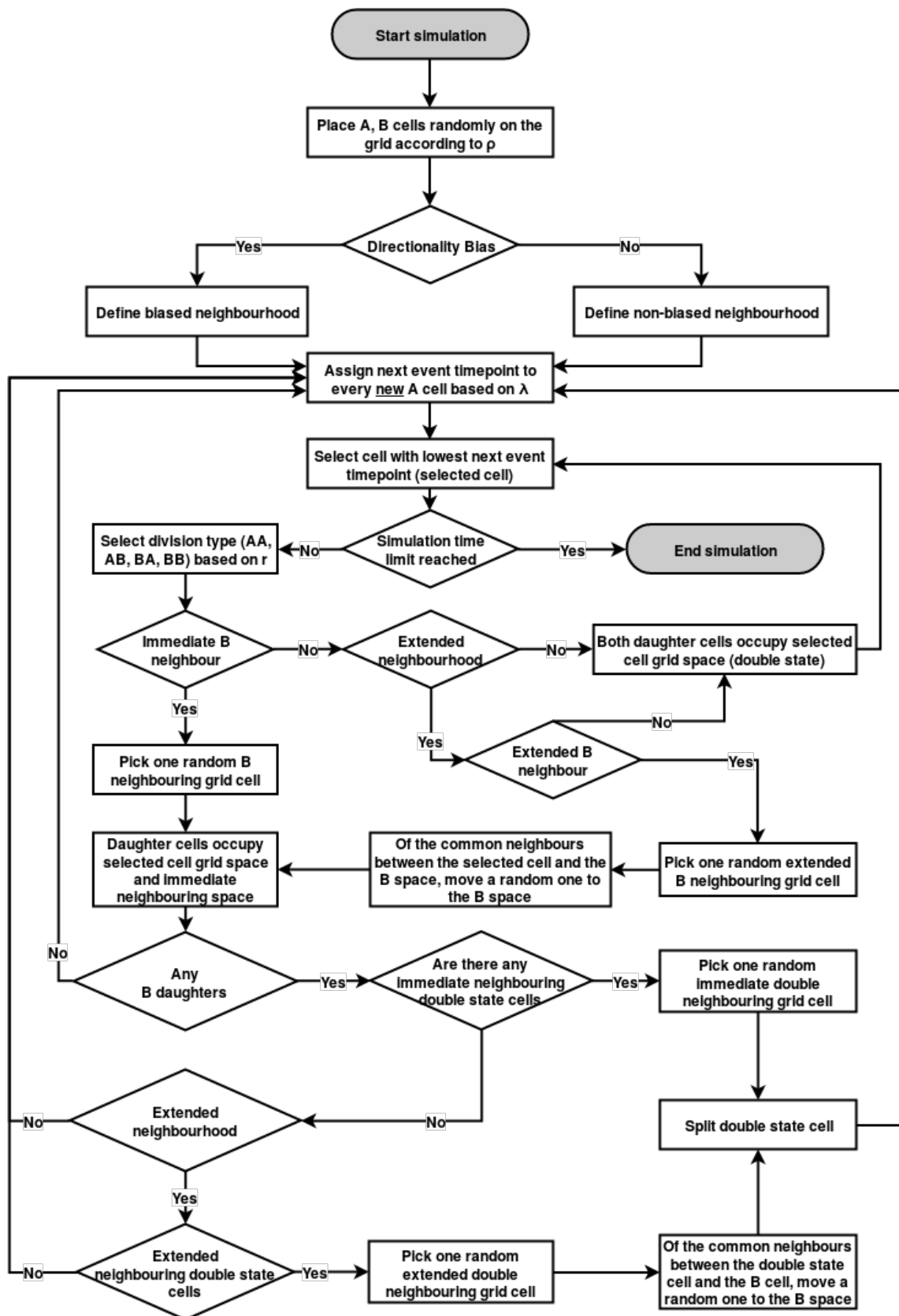
Fig. 4.13 Division driven model model. Detailed flowchart describing the steps of the division driven model.

4.4.2 Division driven spatial simulations of WT oesophageal epithelium

In order to investigate how the division driven hypothesis changes the behaviour of the system, simulations ($n = 100$) were set up and performed in the exact same way as described in Section 4.3, page 80. A visual representation of one simulation at different time points (0, 10, 30, 50 and 70 weeks) is illustrated in Figure 4.14.
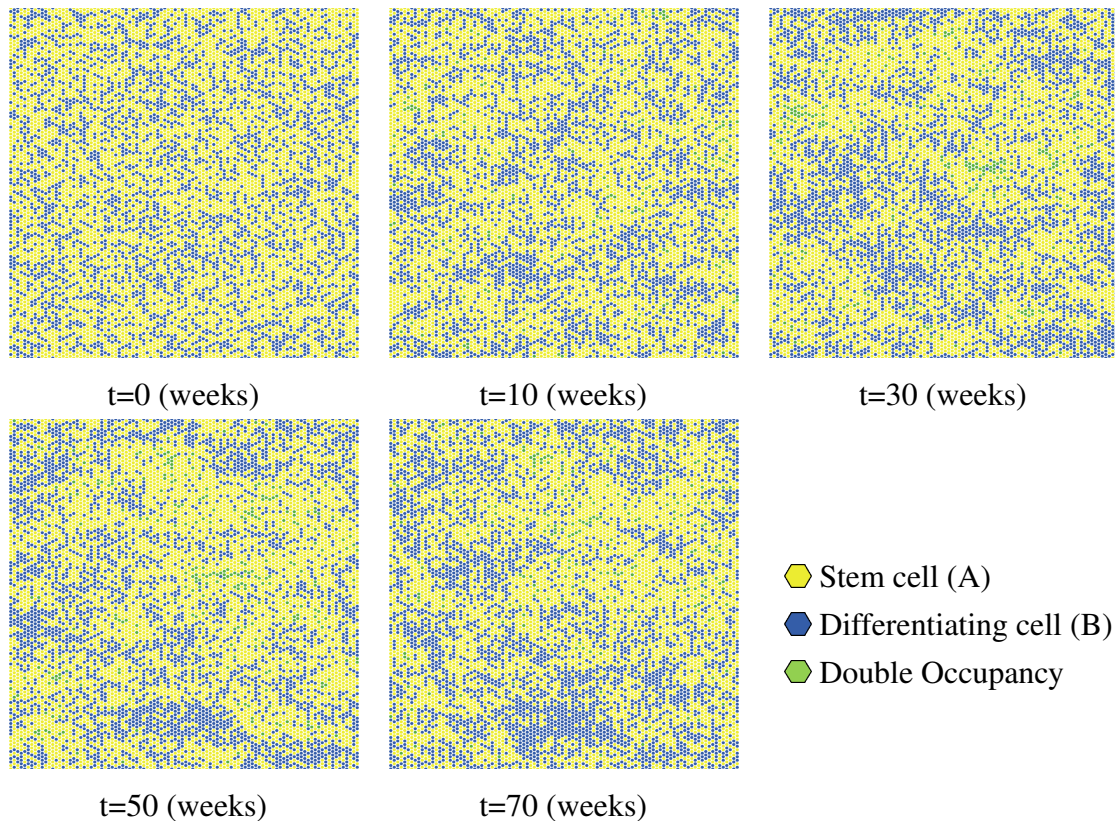
Fig. 4.14 Division driven simulation time lapse. Typical simulation time lapse of the division driven spatial model, simulating cell growth dynamics in mouse oesophageal basal layer at different time points (0, 10, 30, 50 and 70 weeks). Cells are placed on a 2D 100x100 hexagonal lattice. Areas of double occupancies are formed. No empty spaces are observed as a result of the linking between division and stratification. Parameters used: $r = 0.1$, $\rho = 0.65$, $\lambda = 1.9/\text{week}$, from Doupé et al. (2012).

Looking at how the system progresses over time, one can see that, unlike the initial SP spatial implementation, no empty patches are formed, as expected from the linking between division and stratification events (Figures 4.14, 4.15a). Double state cells are still observed, although to a lesser extent (Figures 4.14, 4.15b), in cases when a stem cell divides and there is no adjacent differentiating cell to compensate for this division event. The new SP variant model introduces a tendency of stem and differentiating cells to cluster. This

compartmentalization is evident by the formation of islands of differentiating cells, which progressively become larger (Figure 4.14). Further analysis of the division driven model simulations revealed that an overall homogeneous epithelial basal layer is produced, with the total cell density being nearly constant (Figure 4.15c). Comparison of cell density distribution with lineage tracing microscopy images suggests a significant difference between the two distributions ($***p < 0.001$ by two sample KS test), although the division driven model adopts a much narrower distribution compared to the SP model (Figure 4.15d).

(a)

(b)

(c)

(d)
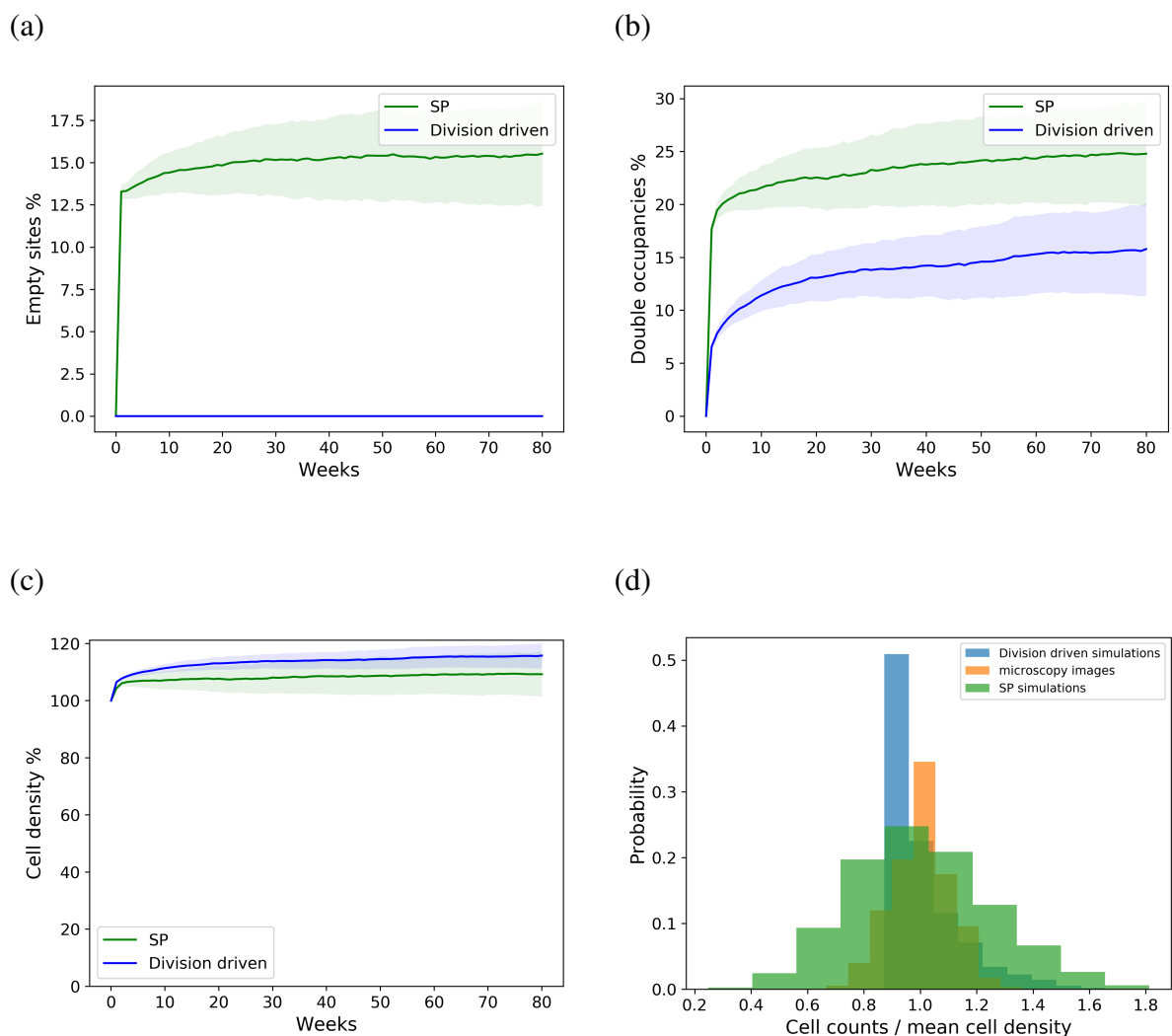
Fig. 4.15 Impacts of division driven spatial model on cell density in mouse oesophagus simulations. a) percentage of empty areas in SP and division driven models, b) percentage of double occupancies in SP and division driven models, c) basal cell density over time. Data correspond to mean values across 100 simulations, d) cell density distributions in confocal microscopy images of murine oesophagus wholemounts, in division driven and SP simulations (t = 52 weeks). Grid was split to 100 segments of 10x10 size. Cell counts were normalized for direct comparison. a,b,c: Shaded areas correspond to SD.

(a)

(b)

(c)

(d)
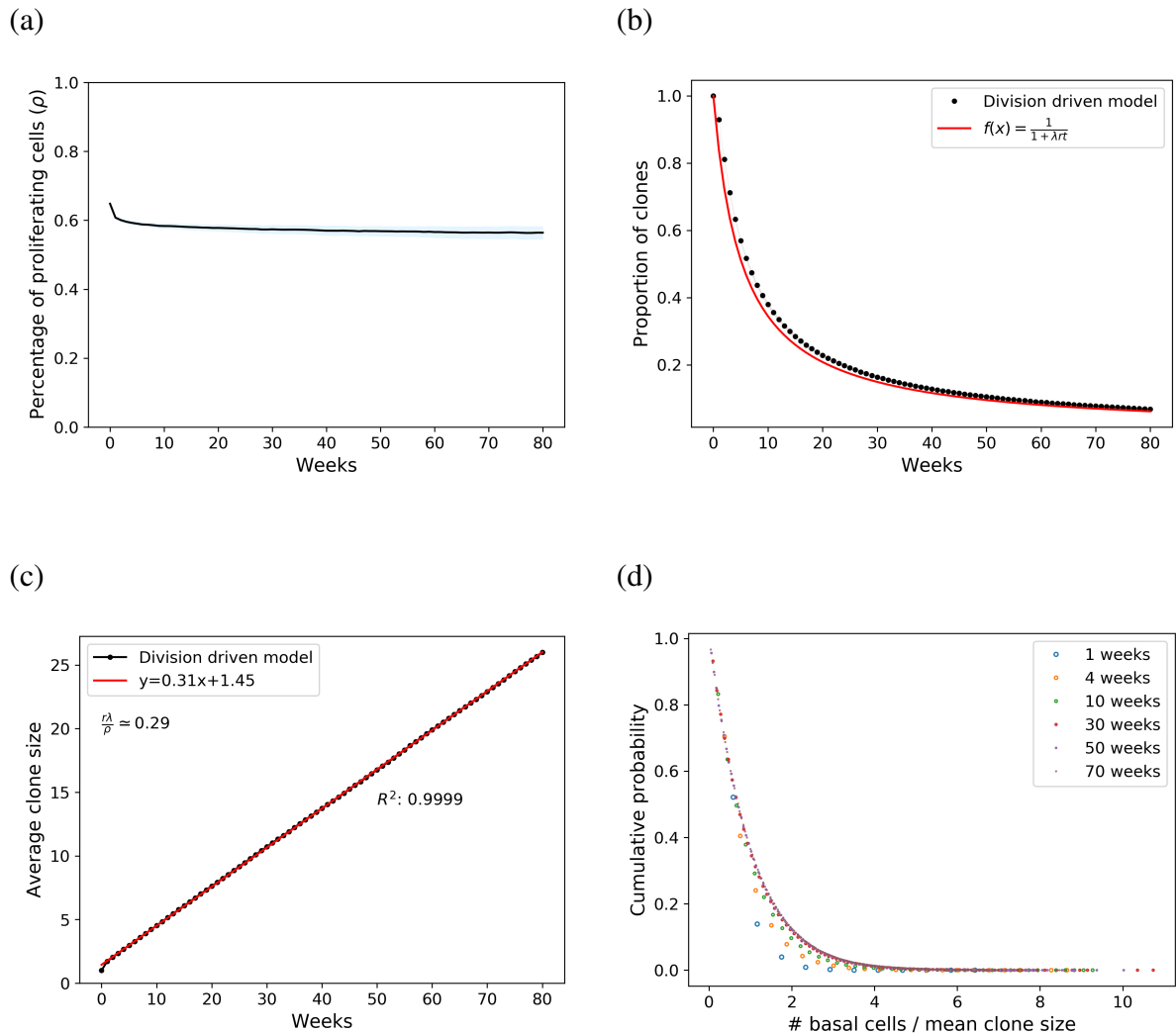
Fig. 4.16 Quantitative analysis of spatial division driven model simulations for the oesophageal dataset. Parameters: $r = 0.1$, $\rho = 0.65$, $\lambda = 1.9$/week, from Doupé et al. (2012). a) stem cell population remains largely constant, b) the number of clones decreases over time, following $\frac{1}{1+\lambda rt}$, c) the average clone size increases over time, with slope $\tau \simeq r\lambda/\rho$, d) clone size distribution exhibits the characteristic scaling over time. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

As demonstrated in Figure 4.16, analysis of cell populations and clone size distributions confirms that the division driven model is also able to recapitulate the signatures of population asymmetry and neutral competition. Similar to the initial model, the balance between cell production and loss maintains a largely constant number of proliferating cells (Figure 4.16a). Furthermore, cell clones stochastically expand or shrink, leading to a drop in clonal population and an increase in the average size of the surviving clones (Figures 4.16b,c) Moreover, the clone size distribution adopts a broader shape over time and exhibits the characteristic scaling behaviour (Figure 4.16d).

Taken together, the fate coupling hypothesis, where division drives local stratification does not modify the SP hallmarks on cell clone population behaviour. More importantly though, the division driven approach does not seem to offer a substantial improvement to the expected dynamics.

4.4.3   Linking stratification to division

In the light of a recent live imaging work on mouse epidermis (Mesa et al., 2018), where it has been suggested that stratification triggers compensatory local cell division, one more alternative version of the single progenitor model of epithelial homeostasis was explored. According to this model hypothesis, division is directly coupled to stratification. That is, division events occur upon a stochastic exit from the basal layer of a nearby differentiating cell. The stratification driven model was implemented as follows:

1. Start by defining a grid of *NxN* sites with A and B cells randomly seeded. The fraction of A cells is calculated from ρ parameter.

2. Draw an exponentially distributed random number with mean $1/\Gamma$ to assign time of next event (stratification) for each B cell.

3. Select B cell with the smallest next time assigned. Current time is updated to the smallest next event time.

4. All neighbouring places are checked for A neighbours. In the case of an existing A neighbour use a random number from a uniform distribution $U \in (0,1)$ to choose the division type to occur by comparing U to division probabilities. One new daughter cell will replace the mother cell and the other will occupy the recently stratified B neighbouring space. If there are no A neighbouring cells, the B cell stratifies leaving an empty space.

5. Repeat steps 3-4 until there are no B cells left or time threshold is reached.

The following flowchart describes the model steps in more detail:
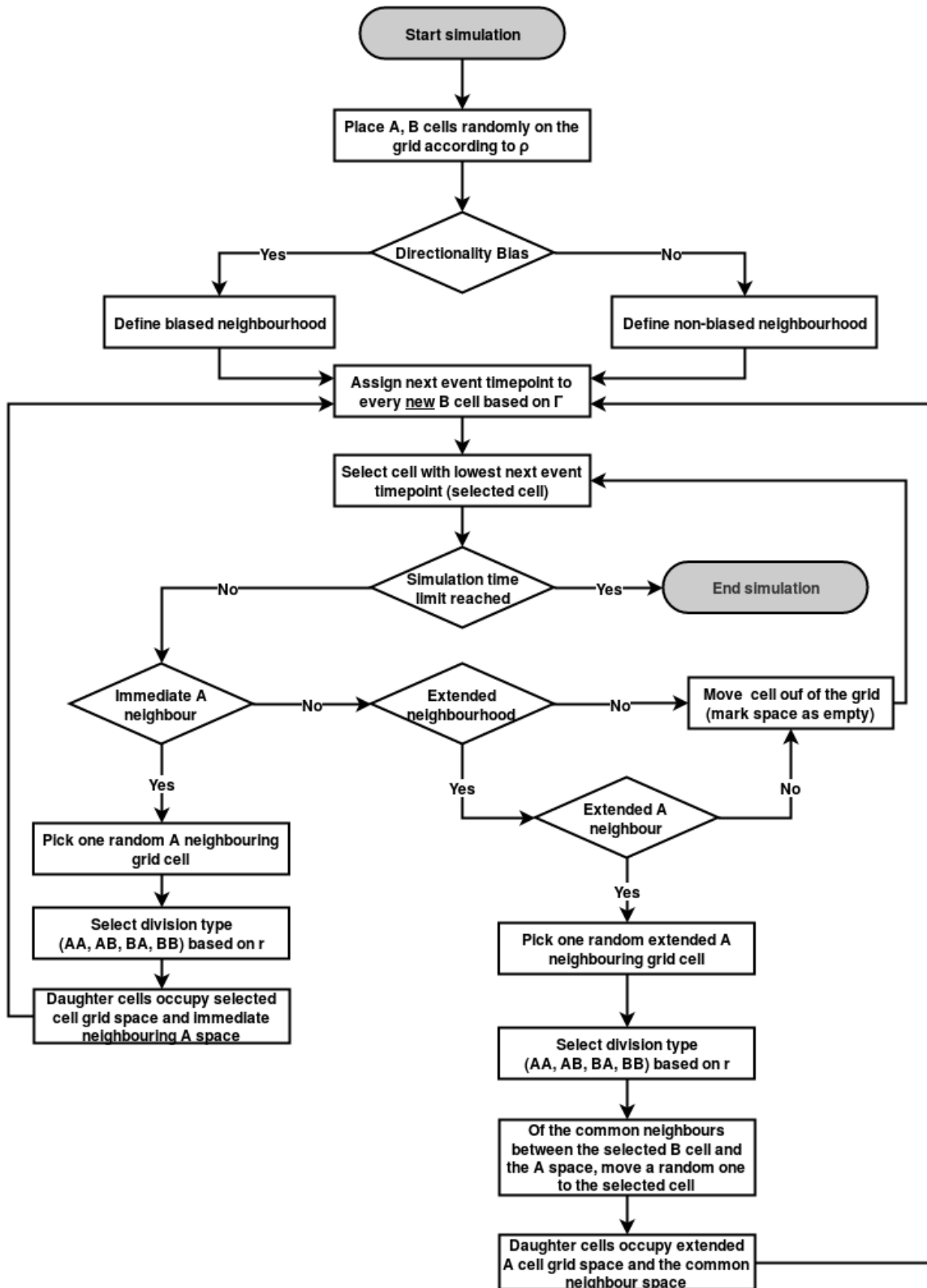
Fig. 4.17 Stratification driven model. Detailed flowchart describing the steps of the stratification driven model.

4.4.4   Stratification driven spatial simulations of WT oesophageal epithelium

The spatial simulations ($n = 100$) of the stratification driven variant were performed in an equivalent way as described in the previous implementations. That is to say, the grid configuration, the duration and the parameter choice were set up to match the ones mentioned in Section 4.3, page 80. A visual representation of the evolution of the simulated tissue at different time points (0, 10, 30, 50 and 70 weeks) is illustrated in Figure 4.18:
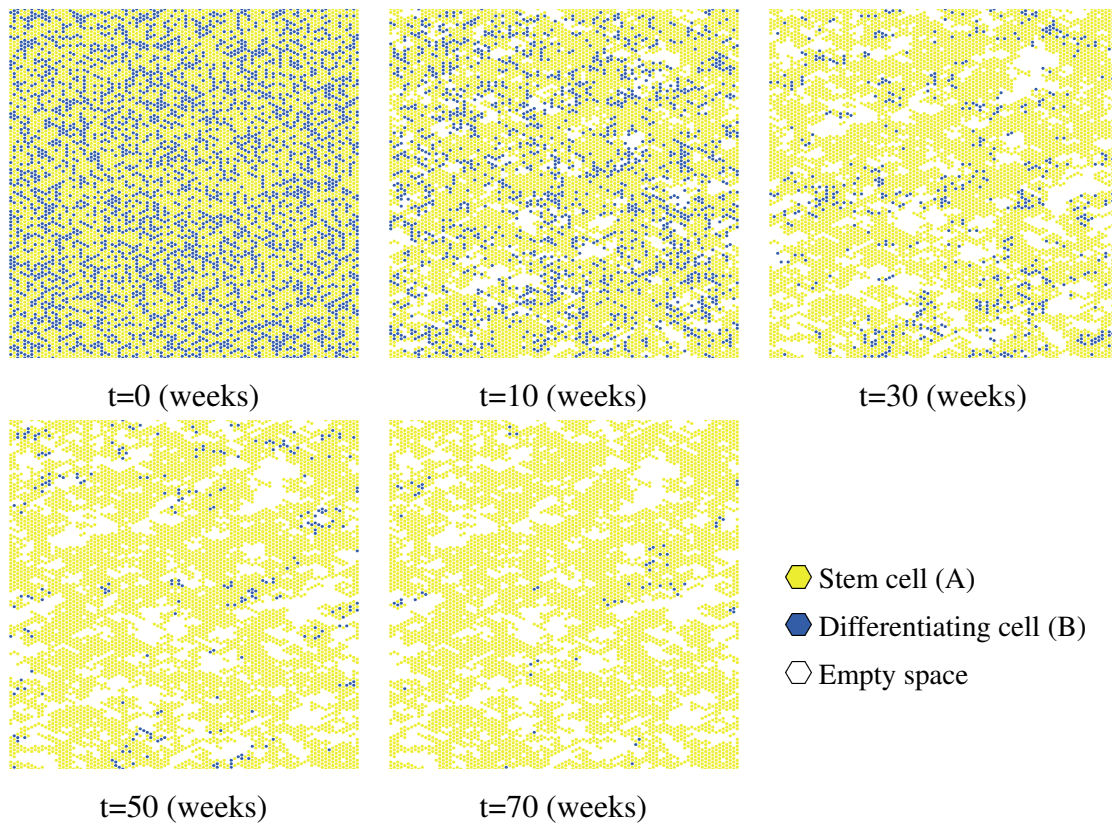
Fig. 4.18 Stratification driven simulation time lapse. Typical simulation time lapse of the stratification driven spatial model, simulating cell growth dynamics in mouse oesophageal basal layer at different time points (0, 10, 30, 50 and 70 weeks). Cells are placed on a 2D 100x100 hexagonal lattice. Gaps are formed which progressively increase in size. The population of differentiating cells drops substantially. Parameters used: $r = 0.1$, $\rho = 0.65$, $\lambda = 1.9$/week, from Doupé et al. (2012).

The simulated spatiotemporal dynamics of the system revealed an insufficiently populated epithelial basal layer. Unlike the previously described model variants (stochastic SP model and division driven model), no islands of double cells are visible when stratification drives division (Figure 4.19a). However, this implementation introduces a non-negligible proportion of vacant sites (Figure 4.19b) within the grid. The increasing incidence of empty regions

leads to a drop in the overall tissue cell density (Figure 4.19c). A quantitative comparison of cell density distribution between simulations and experimental images, as also previously performed on the SP and division driven models, resulted in the same conclusion (Figure 4.19d, $***p < 0.001$ by two sample KS test).

(a)

(b)

(c)

(d)
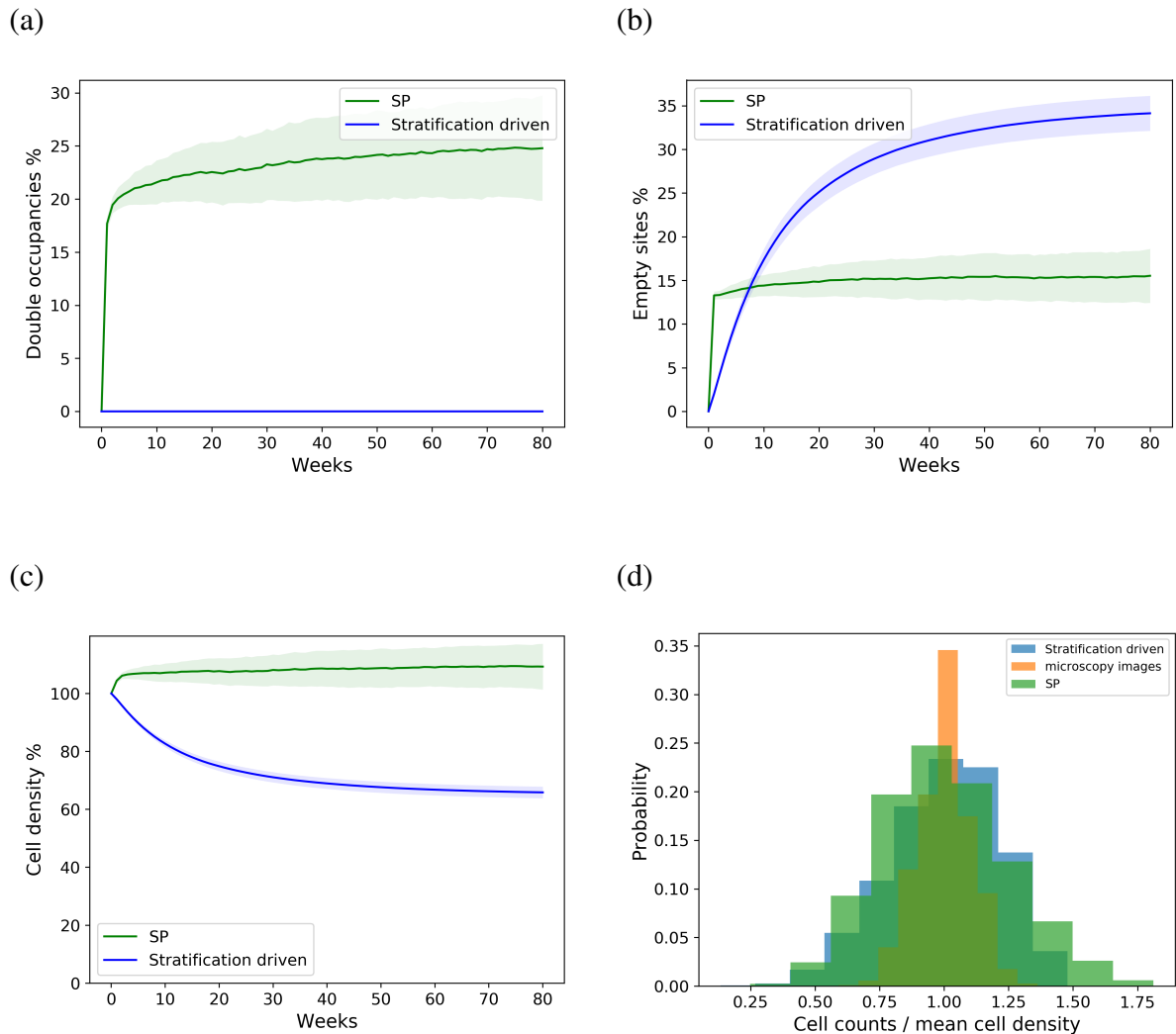
Fig. 4.19 Impacts of stratification driven spatial model on cell density. a) percentage of double occupancies in SP and stratification driven models, b) percentage of empty areas in SP and stratification driven models, c) overall basal cell density drops, d) cell density distribution in confocal microscopy images of murine oesophagus wholemounts, in stratification driven model and SP model (t = 52 weeks). Grid was split to 100 segments of 10x10 size. Cell counts were normalized for direct comparison. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

In addition to the above, analysis of clonal dynamics revealed a series of issues this model variant has in replicating the homeostatic behaviour. The substantial decline in the number of differentiating cells led to an imbalance between cell production and loss, which

is further demonstrated by the increase in progenitor cells' proportion (Figure 4.20a). Clones are lost over time, but not following the expected relationship with r and λ (Figure 4.20b). Furthermore, in contrast to the expected linear growth of average clone size, a sublinear trend is observed and the increasingly broad shape in clone size distributions is not noticeable at later time points. (Figures 4.20c,d).

(a)

(b)

(c)

(d)
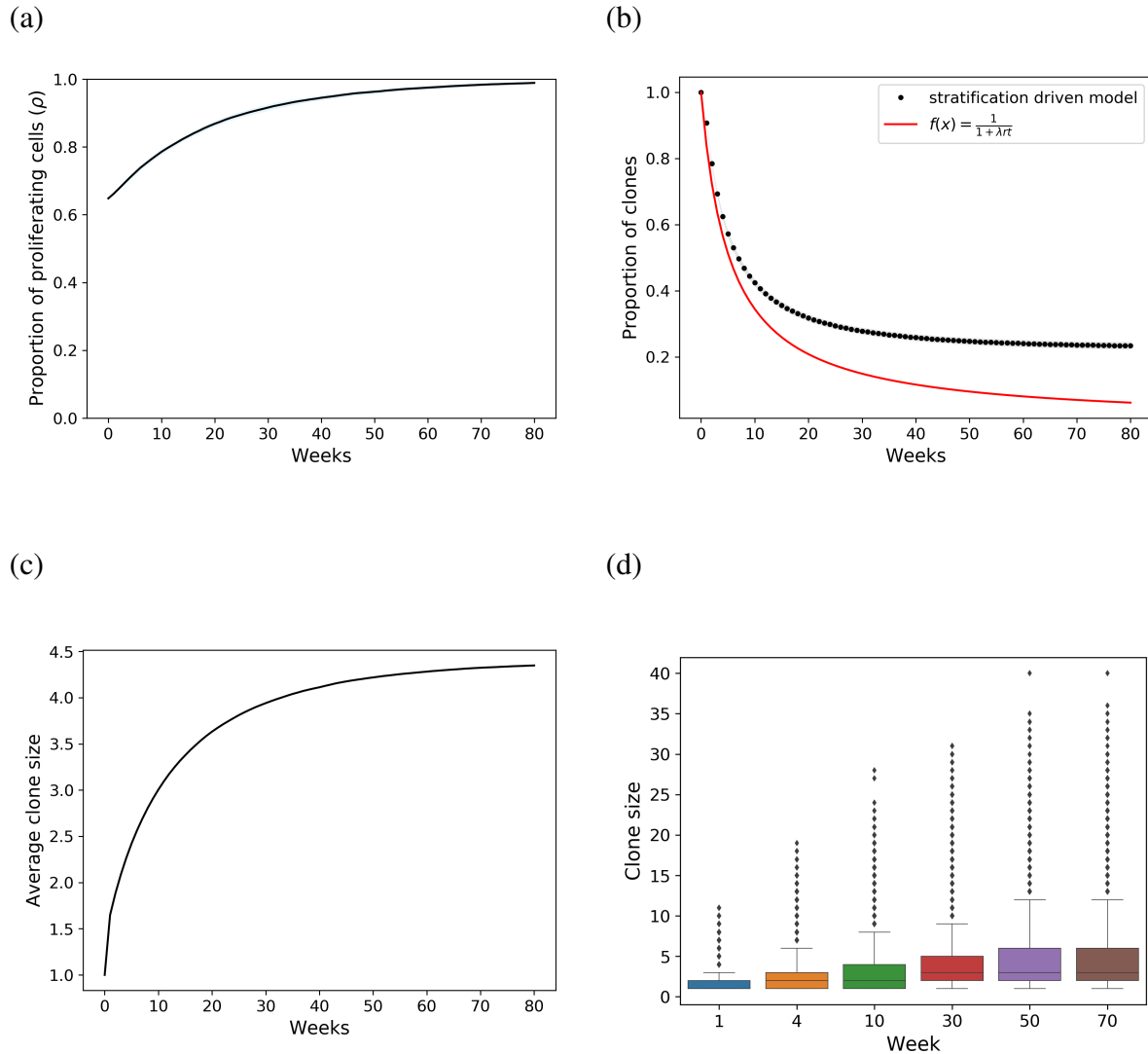
Fig. 4.20 Quantitative analysis of spatial stratification driven model simulations for the oesophageal dataset. Parameters: $r = 0.1$, $\rho = 0.65$, $\lambda = 1.9/\text{week}$, from Doupé et al. (2012). a) proportion of progenitor cells increases, b) the number of clones decreases over time, not following the expected relationship with r and λ, c) the average clone size increases sublinearly over time, d) the increasingly broad shape in clone size distributions is not noticeable at later time points. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

Taken together, using the parameter values that describe the behaviour of the oesophageal progenitor cells, the stratification driven spatial model generates an underpopulated basal layer and fails to reproduce the hallmarks of the SP model of epithelial maintenance. The underlying cause of this problematic behaviour stems from the increasing drop in the population of differentiating cells. As stratification events are indispensable for the time evolution of the system, the lack of differentiating cells renders the simulations starved with patches of gaps and thus unable to continue.

Having identified the cause of the stratification model's pathological behaviour, I sought to assess whether a modified version of it, would be more successful in reproducing the expected homeostatic dynamics. To this end, I implemented a speculative "empty-driven" approach. In the modified version, divisions were still triggered by stratification events with the additional assumption that the placement of a newly generated proliferating cell next to a vacant lattice site would trigger a division of that cell. This approach could reduce the large number of empty sites and assist in supplying differentiating cells to the system. However, it would also indicate that potential continuing placements of newly born cells next to empty sites would cause a chain of consecutive division and stratification events, which is not biologically realistic.

The updated stratification model coupled with the empty driven assumption appeared to fix the pathological behaviour previously introduced by the initial model. As demonstrated in Figure 4.21, the emergence of empty sites is reduced substantially and a homeostatic overall basal cell density is produced. Additionally, quantitative analysis of simulated clonal populations revealed that the SP hallmarks are successfully recapitulated. The proportion of proliferating cells remains constant (Figure 4.22a). Clone survival drops and average clone size increases linearly following the model (Figure 4.22b,c). Moreover, the characteristic scaling behaviour in clone sizes is successfully reproduced (Figure 4.22d). Notably, the possibility of multiple events taking place at the same time leads to an increased number of divisions and stratifications (Figure 4.22e). Furthermore, the replacement of multiple vacant lattice sites as a result of consecutive division events generates a much narrower cell density distribution than that observed experimentally (Figure 4.22f).

Together, these findings demonstrate that the stratification driven model is able to recapitulate the homeostatic behaviour, when an additional empty driven division rule is accounted. Nevertheless, this modification may trigger multiple consecutive divisions in certain cases, which is a biologically unrealistic behaviour. The failure of the implementation of the pure stratification hypothesis raises the question whether the stratification driven approach would

exhibit a more realistic behaviour on alternative parameter sets, or conversely whether the other models would fail under different parameters.

| t=0 (weeks) | t=10 (weeks) | t=30 (weeks) |

| t=50 (weeks) | t=70 (weeks) |

⬡ Stem cell (A)

⬢ Differentiating cell (B)

⬡ Empty space
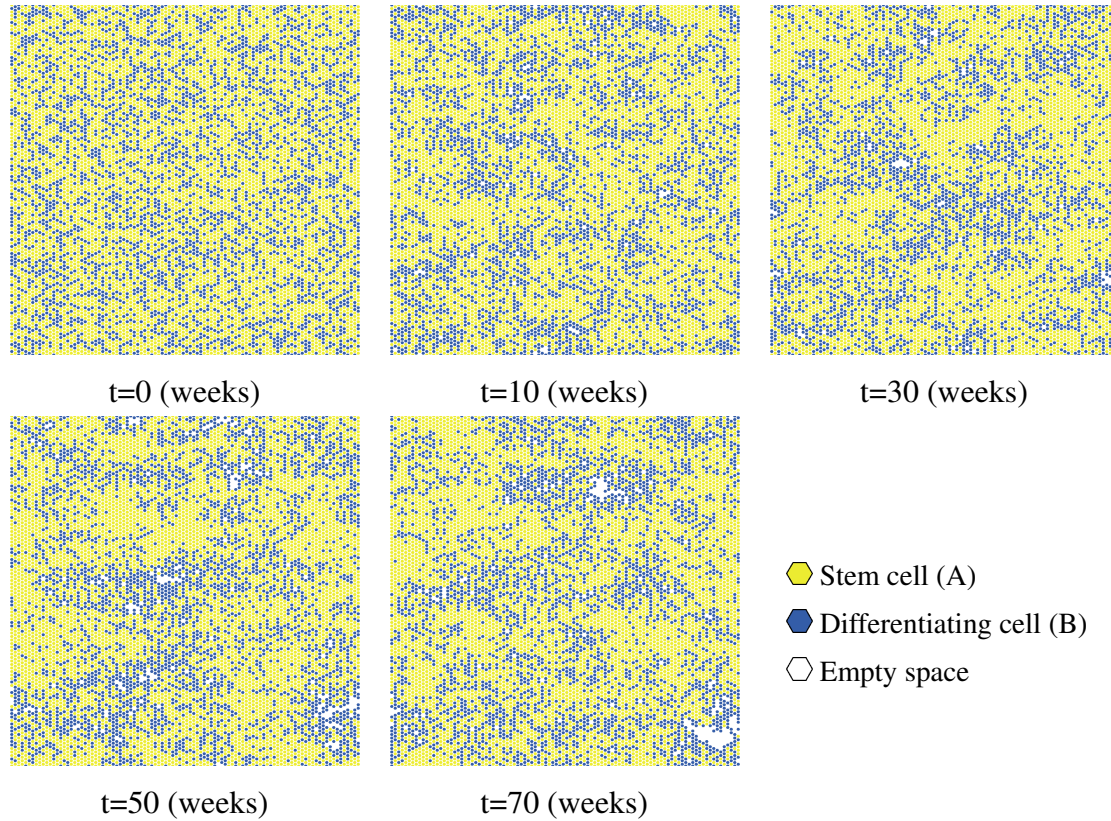
Fig. 4.21 Empty driven simulation time lapse. Typical simulation time lapse of the empty driven spatial model, simulating cell growth dynamics in mouse oesophageal basal layer at different time points (0, 10, 30, 50 and 70 weeks). Cells are placed on a 2D 100x100 hexagonal lattice. The emergence of gaps is reduced substantially and an overall uniform basal cell density is produced. Parameters used: $r = 0.1$, $\rho = 0.65$, $\lambda = 1.9$/week, from Doupé et al. (2012).

(a)

(b)

(c)

(d)

(e)

(f)
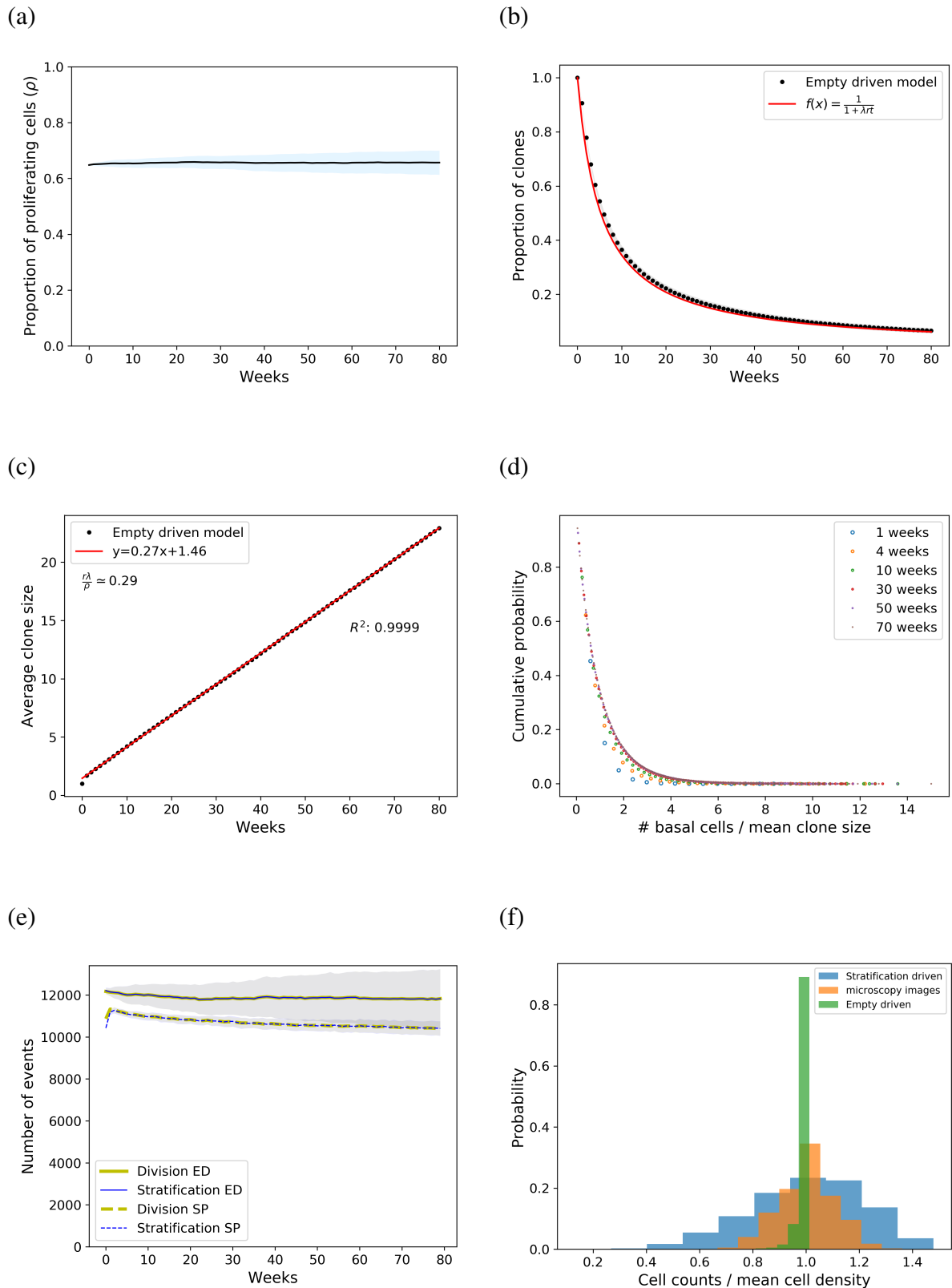
Fig. 4.22 Quantitative analysis of empty driven model simulations on the oesophageal dataset. Parameters: ($r = 0.1$, $\rho = 0.65$, $\lambda = 1.9$/week, from Doupé et al. (2012)). a) The proportion of stem cells remains largely constant, b) the number of clones decreases over time, following $\frac{1}{1+\lambda rt}$, c) the average clone size increases over time, with slope $\tau \simeq r\lambda/\rho$, d) clone size distribution exhibits the characteristic scaling over time, e) the number of division and stratification events in empty driven (ED) model is higher compared to SP model, f) cell density distribution remains highly invariable. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

4.5   Global appropriateness of different models

So far, I have examined the general properties and implications of the three spatial SP variants (independent, division driven, stratification driven) on murine oesophageal tissue dynamics. I found that the expected homeostatic cell population dynamics remain unaltered, excluding the stratification driven model which appears to be the least appropriate for describing the dynamics of this particular tissue. In this section, I will extend the simulations to alternative parameter combinations of oesophagus and other tissue types and will investigate how successfully each model variation reproduces the expected clonal behaviour. For this purpose, I performed computational simulations of each model variant, using as input experimentally inferred parameter values (λ, ρ, r) describing the dynamics of four different murine epithelial tissues (oesophagus, back epidermis, ear epidermis and plantar epidermis) (Table 4.1).

Tissue	Reference	λ(/week)	r	ρ
Oesophagus (OE)	Doupé et al. (2012)	1.9	0.1	0.65
Oesophagus (OE)	Piedrafita et al. (2020)	2.9	0.06	0.56
Back Epidermis (Back)	Murai et al. (2018)	1.16	0.06	0.77
Back Epidermis (Back)	Piedrafita et al. (2020)	1.2	0.04	0.61
Ear Epidermis (Ear)	Doupé et al. (2010)	1.2	0.11	0.28
Ear Epidermis (Ear)	Piedrafita et al. (2020)	1.5	0.04	0.54
Plantar Epidermis (Pl)	Lim et al. (2013)	2.2	0.2	0.8
Plantar Epidermis (Pl)	Piedrafita et al. (2020)	2.0	0.14	0.53
Synthetic dataset (S)	Synthetic dataset (S)	1.5	0.5	0.4
Synthetic dataset (S)	Synthetic dataset (S)	1.5	0.1	0.45
Synthetic dataset (S)	Synthetic dataset (S)	1	0.06	0.35
Synthetic dataset (S)	Synthetic dataset (S)	1.5	0.15	0.7

Table 4.1 SP model parameter sets. Parameter values inferred for progenitor cell behavior in different murine epithelial regions, as derived from quantitative lineage tracing, as well as random sampling parameter values.

As previously shown in the simulations of the oesophageal parameter set, the spatial element introduces density fluctuations, demonstrated by the emergence of either double state cell regions or areas of vacant sites. If the proportion of densely populated or empty regions produced by the spatial model is low, the homeostatic clonal population signatures are not affected. In that case, we can assume that such a model can appropriately depict homeostatic tissue dynamics in two dimensions. The spatially implemented stratification driven model generated a high proportion of empty patches in the lattice, which influenced the clonal population dynamics, thus producing unrealistic emergent properties.

To investigate how the three spatial models perform under different experimentally inferred parameter conditions of epithelial homeostasis, I considered the proportion of double / empty lattice sites as well as the area of the biggest double / empty region as metrics of cell density variability and thus indicators of each model's appropriateness.

Figures 4.23 and 4.24 present the values of the two above mentioned metrics for the experimentally inferred parameter combinations describing different epithelial tissues (Table 4.1), at 10 and 50 weeks respectively. A set of random sampled parameter sets were also used (denoted as "S") in an effort to make the parameter space more evenly spaced.

During the 10th week, the spatial models are largely able to achieve homeostatic conditions for each tissue type, apart from plantar epidermal parameters (Lim et al., 2013), as both the SP and division driven models produce an increased proportion of high density areas given the short simulation period. More specifically, whilst the percentages of double sites and largest empty island had similar values across all other parameter sets and did not exceed 20% and 10% respectively, simulations of Lim et al. (2013) plantar parameters produced a much higher percentage of such regions ($\sim$35% and 20% respectively).

At 50 weeks, it can be observed that the SP and division driven models can sufficiently reproduce a relatively low proportion of high density areas for most of the parameter sets, with the division driven model achieving a lower percentage of crowding regions. The plantar epidermis (Lim et al., 2013) appears to have a high percentage of double areas (>35%), therefore the SP and division driven regimes are less successful in reproducing tissue maintenance behaviour for this parameter set. The stratification driven hypothesis seems to be the least appropriate model for most of the tested parameters in oesophagus, ear and plantar epidermis (Piedrafita et al., 2020), as it generates high incidence of gaps in the simulated tissue. However, it is worth mentioning, that the gap percentage for back (Murai et al., 2018, Piedrafita et al., 2020) and plantar (Lim et al., 2013) epidermis is lower. From the three, the parameter set describing back epidermis homeostasis (Murai et al., 2018) achieves the lowest empty site percentage (<10%). This would suggest that the stratification driven hypothesis could be a more appropriate model describing tissue homeostasis in this tissue type. The stratification driven hypothesis has been also proposed by Mesa et al. (2018) as a homeostatic mode of tissue maintenance, following live imaging in mouse plantar epidermis.

Driven by this, I further examined whether the stratification driven model can successfully reproduce the homeostatic signatures of cell population dynamics using the back epidermis (Murai et al., 2018, Piedrafita et al., 2020) and plantar (Lim et al., 2013) data sets. Despite the percentage of vacant sites being sufficiently low, neither back skin (Piedrafita et al., 2020) nor plantar (Lim et al., 2013) data sets were able to reproduce the key signatures of homeostasis.
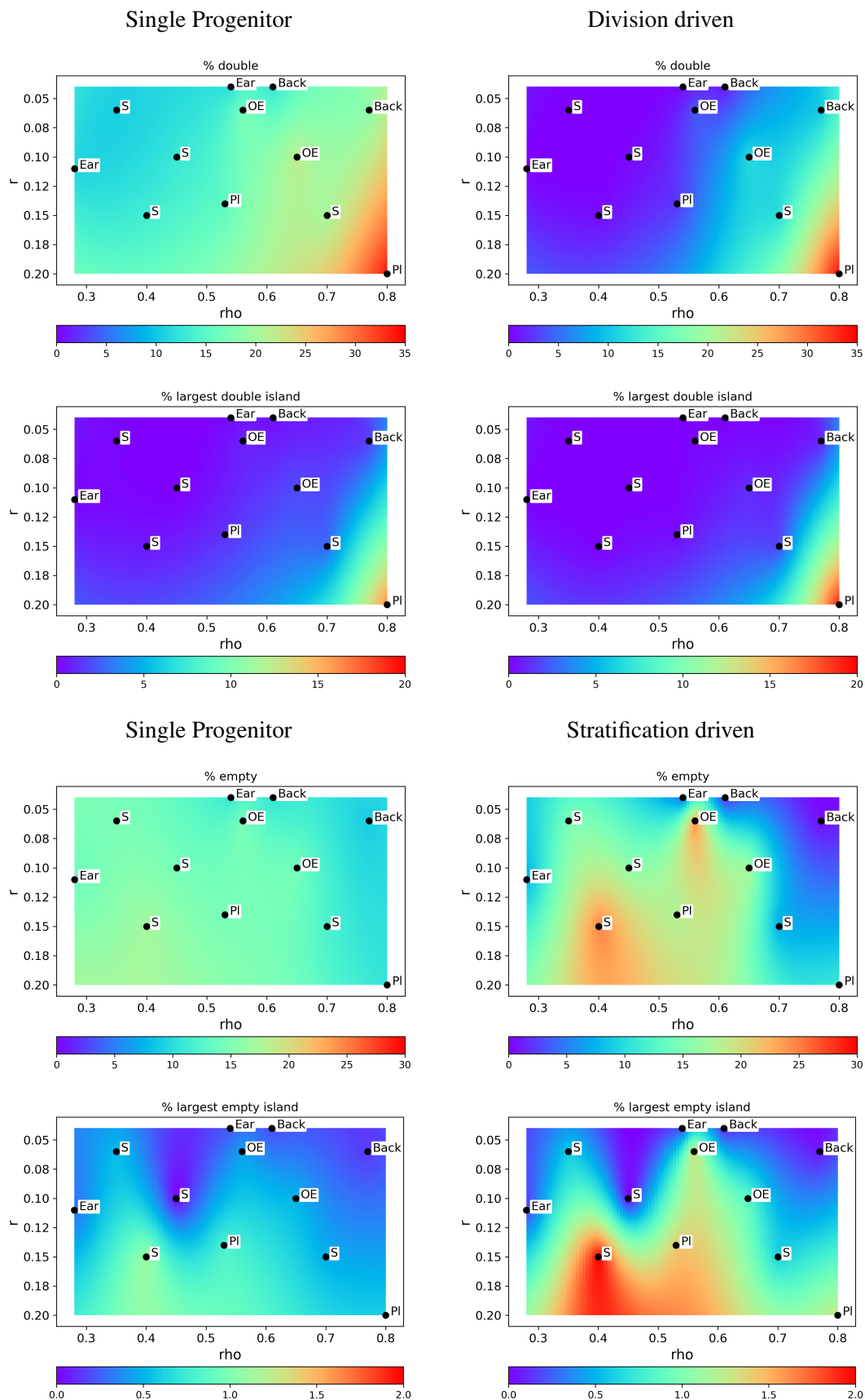
Fig. 4.23 Model comparison on various datasets (t = 10 weeks). Comparison of density variability metrics (% double / empty lattice sites, % largest double / empty area) in the three spatial models under a set of experimentally inferred and randomly sampled parameter combinations describing different epithelial tissue regions. Dots correspond to parameter sets described in Table 4.1.
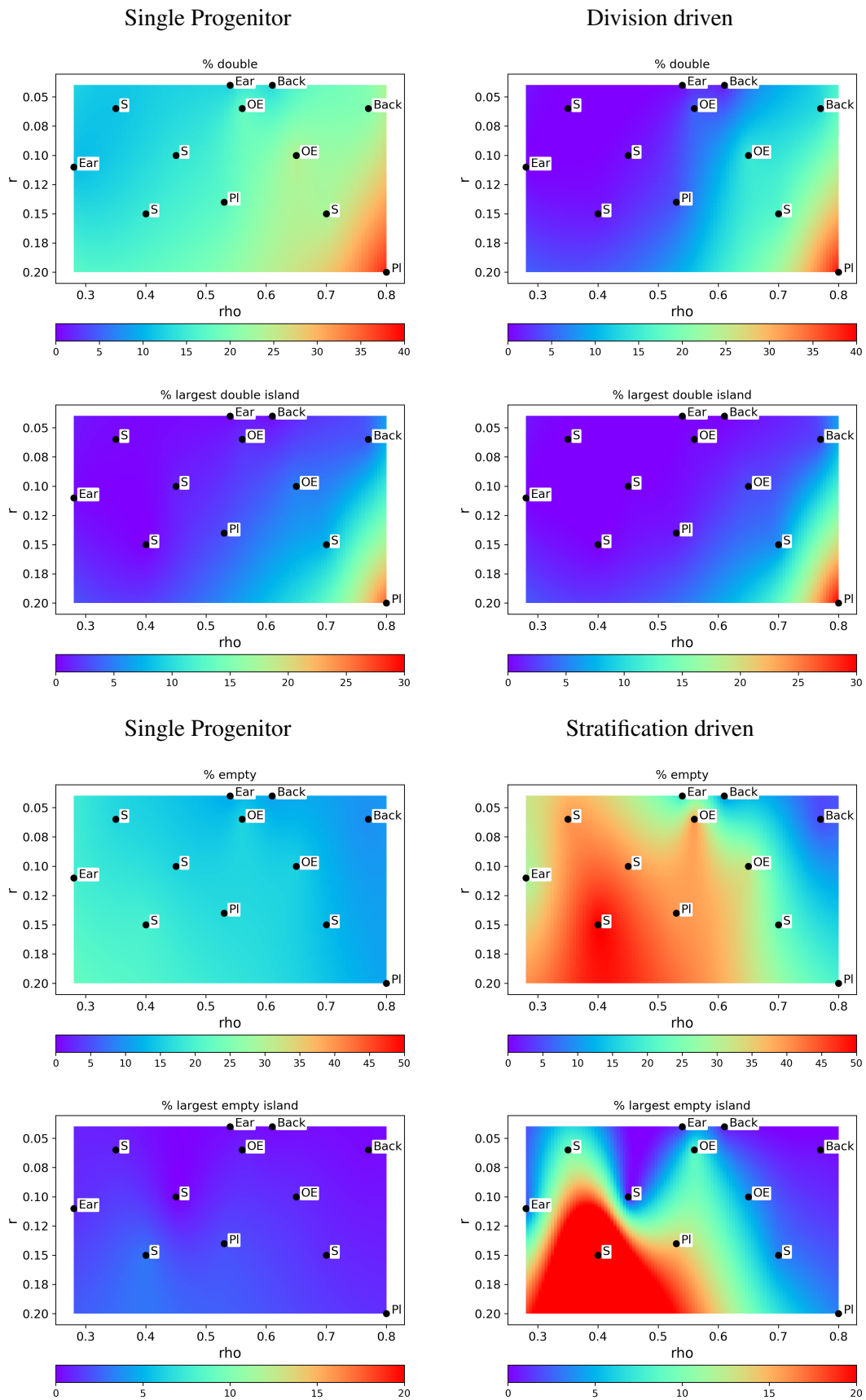
Fig. 4.24 Model comparison on various datasets (t = 50 weeks). Comparison of density variability metrics (% double / empty lattice sites, % largest double / empty area) in the three spatial models under a set of experimentally inferred and randomly sampled parameter combinations describing different epithelial tissue regions. Dots correspond to parameter sets described in Table 4.1.

(a)

(b)

(c)

(d)
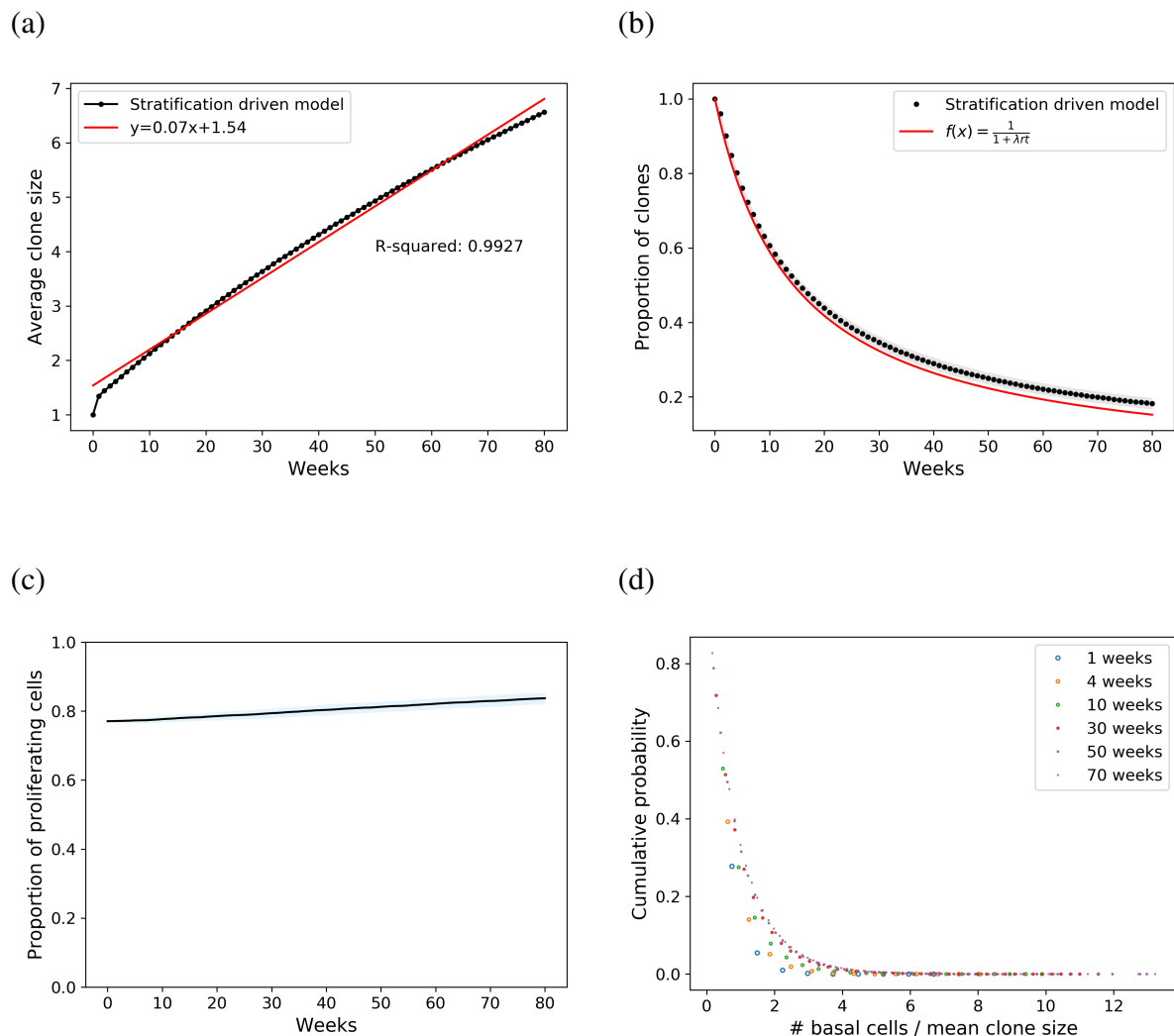
Fig. 4.25 Quantitative analysis of spatial stratification driven model simulations for the back epidermis dataset. Parameters: $r = 0.06$, $\rho = 0.77$, $\lambda = 1.16$/week, from Murai et al. (2018). a) the average clone size increases over time, b) the number of surviving clones decreases over time, c) proportion of progenitor cells, d) clone size distribution scales with time. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

Interestingly however, quantitative clonal analysis of the stratification driven model with back skin parameters (Murai et al., 2018) showed a completely different picture to what was previously observed in stratification driven simulations. In contrast to a sublinear trend in average clone size (Figure 4.20c), a nearly linear growth was observed with slope value being close but not matching the expected $r\lambda/\rho$ (0.07 instead of 0.09, Figure 4.25a).

The number of clones dropped, nearly matching $1/(1 + \lambda rt)$ (Figure 4.25b). Moreover, the proportion of proliferating cells remained largely constant, highlighting no substantial decline in post mitotic cell population (Figure 4.25c), and the distribution of clone sizes

appeared to follow the characteristic scaling behaviour (Figure 4.25d). Together, all the above findings suggest that the stratification driven hypothesis could be a more appropriate model in describing the dynamics of certain data sets.

4.6   Discussion

To summarize, the previously discarded spatial information provided by the lineage tracing experiments can be used to give new insights into the coordinated behaviour of epithelial stem cells in the context of a tightly packed tissue. Therefore, in this chapter I sought to explore epithelial stem cell growth dynamics in homeostasis when tissue organisation is explicitly included. To this end, a CA was used to implement the SP model in two dimensional space.

Simulations of the spatial SP model were initially set up to mimic the dynamics of the WT mouse oesophageal basal layer. The simulation outputs revealed that the spatial model is able to reproduce the characteristic features on clonal population dynamics dictated by the SP paradigm. Interestingly, the explicit accounting of space introduces local, transient density fluctuations, demonstrated by the emergence of either double state cell regions or areas of vacant sites.

The above outlined observations reflect the stochastic, cell intrinsic control of cell fate imposed by the SP model. Cell division and exit of differentiated cells from the basal layer, albeit two balanced events at the cell population level, are not directly linked. In a real tissue however, local spatial coordination mechanisms should exist to ensure tissue level cellular equilibrium (O'Brien and Bilder, 2013). This would imply that the otherwise successful stochastic model of epithelial homeostasis in the zero dimension might need the incorporation of cell communication feedback mechanisms to fully capture tissue dynamics in the two-dimensional space.

Driven by this, the SP model was modified to consider spatial influences on stem cell stochastic behaviour. The model variant would assume that division and stratification events are no longer two independent processes, but they are directly coupled. Two different scenarios were examined: division drives local stratification and stratification drives local division, two concepts also supported in the literature (Eisenhoffer et al., 2012, Marinari et al., 2012, Miroshnikova et al., 2018, Mesa et al., 2018). Simulations of the two SP variants on oesophagus revealed that the division driven hypothesis is capable of reproducing the SP hallmarks, whereas the stratification driven approach produces an insufficiently populated epithelial basal layer containing large areas of vacant sites.

The three spatial models were also tested on alternative epithelial tissues (back, ear, plantar epidermis). The SP and division driven models were able to describe sufficiently most of the data. Interestingly, the stratification driven hypothesis consistently failed to recover homeostatic conditions for all parameter sets except back skin (parameters taken from Murai et al. (2018)). This would suggest that this model could be more appropriate for

describing the dynamics of particular datasets. Notably, both division driven and stratification driven (when successful) models did not offer a substantial improvement to the expected homeostatic dynamics.

Taken together, despite the emergent property of local cell density heterogeneity, I demonstrated that the SP model continues to match experimental data when neutral clonal competition in space is explicitly considered. The spatial simulations successfully reproduce the fundamental SP signatures on cell populations and therefore the properties of the system are not affected by the lattice geometry. Thus the spatial SP could be considered an appropriate model to study the neutral growth dynamics of epithelial stem cells during homeostasis.

Chapter 5

Non neutral growth is influenced by spatial constraints and feedbacks

Abstract

The accumulation of non-neutral mutations in cells may result in the formation of persistent clones which could potentially lead to cancer. In chapter 4, I demonstrated that spatial constraints do not alter wild-type, neutral epithelial stem cell clonal dynamics. In the current chapter, I set out to explore the effect of spatial structure on tissue growth and clonal competition when mutations are non-neutral. For this purpose, the spatial single progenitor model of epithelial stem cell dynamics was extended to simulate the behaviour of a mutant tissue. I find that in order to recapitulate non-neutral mutant dynamics whilst maintaining tissue turnover the spatial models need to take account of feedbacks between neighbouring cells in the tissue. More specifically, I investigate the growth of p53 mutants in mouse skin epidermis and define an appropriate feedback mechanism which recapitulates their behaviour. This mechanism was successfully validated on p53 mutant clonal data from a different tissue, mouse oesophageal epithelium. I further demonstrate that the growth dynamics of Notch pathway mutants in mouse oesophagus is described by different feedback rules. These findings are able to mechanistically distinguish the behaviour of different mutations. I then explore how these two mutant types compete in space.

5.1   Introduction

The growth and competition of cells in epithelial tissues plays an important role in both tissue homeostasis and the robustness to pre-cancer mutation. In such crowded environments, cell populations constantly compete for space in a controlled manner that minimizes redundant or insufficient cell production and sustains homeostasis. This may be accomplished by mechanical interactions, e.g. through the mechanically activated ion channel Piezo1 (Gudipaty and Rosenblatt, 2017, Eisenhoffer et al., 2012), and molecular signalling, e.g. through Notch signalling pathway (Guruharsha et al., 2012), within the tissue. It is worth noting that recent theoretical studies highlight the role of mechanical cell competition in tissue development and homeostasis (Lee and Morishita, 2017, Pan et al., 2016, Bielmeier et al., 2016). Whilst wild-type cells compete neutrally for dominance in the un-mutated tissue, naturally occurring mutations in individual cells may lend them a fitness advantage that can allow tissue takeover. Thus, non neutral competition between cells with different fitness levels leads to the dominance of the fitter population ("winners") at the expense of the less fit population ("losers"). This is a common mechanism believed to be used by oncogenic mutations to colonize the tissue and potentially drive preneoplasia and tumour formation.

Mutations in the tumour suppressor gene p53 and Notch pathway have been shown to exhibit non neutral growth dynamics in epithelial tissues. An increasing body of work suggests that p53 mutations and inactivation of Notch receptors are highly frequent in squamous epithelial cancers (Alcolea et al., 2014, Agrawal et al., 2011, Stransky et al., 2011). Furthermore, recent studies report high incidence of p53 and Notch1 mutated genes in normal human skin and oesophagus (Martincorena et al., 2015, 2018, Yokoyama et al., 2019). This highlights the importance of studying the process of accumulation and interaction of these mutations within tissues in order to understand tumourigenesis. Considering the above findings, I focused on two lineage tracing experiments which used mouse models to investigate the growth of p53 and Notch signalling pathway mutations in stratified squamous epithelial tissues.

Murai et al. studied the fate of epidermal epithelial cells carrying a heterozygous p53 gain-of-function mutation $p53^{R245W}$, the mouse equivalent of the frequently detected $p53^{R248W}$ in both normal and cancerous human epidermis. By inducing a $p53^{R245W}$ ($p53^{*/wt}$) mutation in transgenic mouse epidermal progenitor cells, the growth dynamics of mutants in a background of their wild-type counterparts was tracked.

Up to 24 weeks post induction, the p53 mutant population outcompeted normal epidermal cells, clearly indicating their competitive advantage. Beyond that transient period of increased

mutant clonal growth, the expansion rate of $p53^{*/wt}$ cells weakened considerably and they eventually reached a 30% proportion of the basal layer by 15 months. Based on 5-ethynyl-2'-deoxyuridine(EdU) data, a marker that stains basal cells in S phase of the cell cycle (i.e. the phase during which DNA replication occurs), the restricted mutant growth at later time points could not be attributed to a decrease in mutant division rate. Along with the slowing of mutant population expansion at 24 weeks, a roughly 10% increase in basal layer cell density was observed, without influencing epidermal functional integrity.

Together, those findings suggested that $p53^{*/wt}$ epidermal progenitor cells adjust their fate to respond to alterations in their cellular environment. That is to say, the crowding induced stress in the basal layer as a result of mutant non-neutral growth may force the return toward balanced dynamics (Le et al., 2016, Roshan et al., 2016).

Alcolea et al. studied the effects of Notch pathway inhibition in mouse oesophageal epithelial cells. To do this, they blocked the transcription of Notch receptor's intracellular domain using mice expressing a negative form of Mastermind-like 1 (Maml1) protein. The dynamics of the mutated oesophageal epithelium was followed for 1 year. Over the first 3 months post induction Maml knock outs expanded rapidly at the expense of their wild-type neighbours. At later time points, mutant growth rate slowed down as soon as Maml clones started merging. However, the proportion of mutant cells kept increasing until the entire epithelium was colonised. Despite the overall epithelium being replaced by Notch mutants, no tumours were formed.

Quantitative clonal analysis suggested that Notch mutants did not undergo divisions producing two differentiating daughters (i.e. their symmetric differentiation probability was turned off) and they also promoted the differentiation of the adjacent wild-type cells. Moreover, they appeared to have an increased division rate and decreased stratification rate. These observations lend Maml mutants an apparent competitive advantage and explain their ability to spread across the tissue. Strikingly, once tissue take over completed, the imbalance in mutant fate reverted towards normal and a new stationary state with a faster turnover was established.

Collectively, the studied growth of p53 and Maml mutations in mouse epithelia has been shown to have distinct properties. Whilst both mutant types switch their fate towards the production of persistent clones, their ability to colonize the tissue differs. Studying the causes of these distinct mutant growth behaviours would assist to distinguish between less and more aggressive expansion mechanisms.

5.2   Spatial SP model of non-neutral competition

The study of p53 and Maml mutant clone dynamics in mouse epithelia has indicated that the mutant progenitor cells present a fitness advantage over their wild-type (WT) counterparts. Different potential scenarios can be considered to explain the observed mutant tissue takeover:

- The division rate of mutant cells is faster than that of WT cells.

- The stratification rate of mutant cells is decreased.

- Mutant cells tend to produce a higher proportion of proliferating cells compared to differentiating cells.

The first two hypotheses do not alter the balanced division outcomes in SP behaviour and could be modelled by a SP model with altered λ and Γ parameters, leading to a transient period of expansion before a new stationary state is restored, as observed experimentally. The third hypothesis, however, biases the division probability outcomes in such a way that divisions resulting in pairs of stem cells become more likely than divisions giving differentiated daughters.

There is no indication of p53 mutant cells dividing at different rates to WT, according to EdU data (Murai et al., 2018). As already mentioned, Maml clones were found to proliferate and stratify at a higher and lower rate respectively. However, quantitative clonal analysis in both p53 and Maml mutant clones revealed that a bias towards the production of proliferating progeny is able to explain the observed mutant behaviour. Such bias results in a gradual expansion in the proliferating population over time as there are fewer chances for the mutant clones to be lost by differentiation. Thus, mutant clonal expansion appears to be consistent with a SP model including a cell fate imbalance (Klein et al., 2010, Frede et al., 2016):

$$
A \xrightarrow{\lambda}
\begin{cases}
AA & r(1+\delta) \\
AB & 1-2r \\
BB & r(1-\delta)
\end{cases}
\tag{5.1}
$$

$$
B \xrightarrow{\Gamma} \varnothing,
$$

where δ denotes the tilt in cell fate. Therefore, $\delta = 0$ corresponds to homeostasis, $\delta = 1$ implies absence of symmetric differentiation, leading to persistence and $\delta = -1$ implies absence of symmetric division, leading to extrusion.

To simulate mutant clonal dynamics in two-dimensional space, I used the spatial SP model with fate imbalance, considering both wild type and mutant epithelial cells. The choice of the grid architecture, the neighbourhood, the number of states and spatial rules of the CA model were implemented as described in Section 4.2, page 73. However, a new mutation-status property was introduced to distinguish mutant cells to wild type cells. In the case of p53 cells, mutants have the exact same properties as the wild type cells and the only property that distinguishes the two cell populations is that p53 mutant cells have an innate bias towards the production of proliferating cells. In the case of Maml cells, mutants will also have distinct λ, Γ and r values, as inferred by Alcolea et al. (2014). Considering the above, the spatial SP model, initially described in equation (4.1), page 76 is modified as follows, in order to accommodate the mutant cell population:

$$A \varnothing \xrightarrow{\lambda} \begin{cases} AA & r(1+\delta) \\ AB & \frac{1}{2} - r \\ BA & \frac{1}{2} - r \\ BB & r(1-\delta) \end{cases}$$

$$A X \xrightarrow{\lambda} \begin{cases} D_{AA} & r(1+\delta) \\ D_{AB} & \frac{1}{2} - r \\ D_{BA} & \frac{1}{2} - r \\ D_{BB} & r(1-\delta) \end{cases} \tag{5.2}$$

$$B \xrightarrow{\Gamma} \varnothing,$$

where $A \varnothing$ denotes a type A cell neighbouring a vacant lattice site and AX denotes a type A cell neighbouring either a type A or type B cell, thus indicating that there is no neighbouring empty space. D_{AA}, D_{AB}, D_{BA}, D_{BB} correspond to double occupancies.

Figure 5.1 represents a schematic description of the spatial rules of the non neutral CA model.
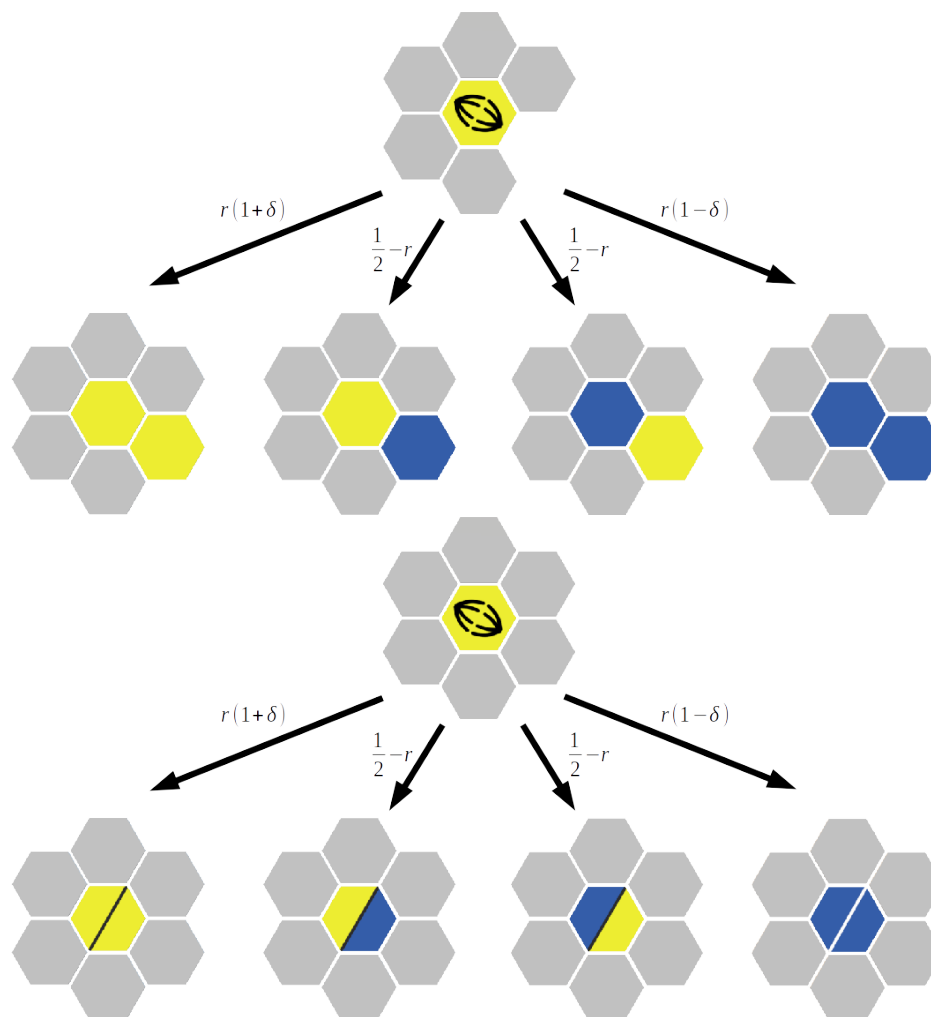
Fig. 5.1 Rules of the non-neutral spatial SP model. Illustration of the spatial SP model including a bias in cell fate. Proliferating cells in yellow and differentiating cells in blue. Proliferating cells undergo a division type which is decided by the probability of symmetric divisions, r and the bias δ. In case of wild type cells, $\delta = 0$ and division probabilities remain unchanged. In case of mutant cells, $0 < \delta \leq 1$, hence symmetric division (AA) is favoured. If a vacant site exists (top), one daughter cell occupies the mother cell's space and the second the neighbouring empty space. If there is no empty space in the immediate neighbour (bottom), the two daughters occupy the mother cell's space, thus creating a double cell occupancy. Double state cells are released once a neighbouring lattice site becomes available.

To simulate the time evolution of the system, the Markovian stochastic simulation algorithm, initially described in 4.2 was modified as follows:

1. Start by defining a grid of *NxN* sites with A and B cells randomly seeded. The fraction of A cells is calculated ρ parameter.

2. Insert randomly mutant cells of type A.

3. For every cell on each lattice site, draw a random number from an exponential distribution with mean $1/\lambda$ or $1/\Gamma$ to assign time of next event (division or stratification) for A and B cells respectively.

4. Select cell with the smallest next event time assigned. Current time is updated to the smallest next event time.

5. If an A cell is selected, use a random number from a uniform distribution $U \in (0,1)$ to choose the division type to occur by comparing U to division probabilities. Symmetric division probabilities for mutant cells are biased according to δ (5.2). Assign the division type as a next event for the selected cell. If a B cell is selected, assign stratification as a next event for the selected cell.

6. If next event is division, all neighbouring places are checked for empty space. In the case of an existing neighbouring space, one new daughter cell will replace the mother cell and the other will occupy the empty neighbouring space. If there is no empty neighbouring space available then both will remain at the mother cell's space (creating a "double state" cell), until a neighbouring space is released. If stratification is the next event, B cell stratifies, leaving an empty space, which allows a pair of potential neighbouring "double state" daughters to be released.

7. Repeat steps 4-7 until there are no A or B cells left or time threshold is reached.

5.3 Logistic growth quantitatively explains takeover of p53 mutant cells

To address the question of how the spatial system responds to the induction of p53 mutations, computational simulations of the non neutral SP CA model were performed. The spatial SP model with biased proliferation probabilities was used to explore whether the above experimental observations can be reproduced and explain the causes of p53 mutant growth slowness. The simulations were set up and performed as described in Section 4.3, page 80. Initial simulations of the p53 clone takeover modified a set of wild-type cells to have an imbalance in symmetric cell fates. To replicate the lineage tracing experiment, 1% of p53 mutant cells, matching the reported mutant induction (Murai et al., 2018), were also introduced at the beginning of each simulation. The following parameter values from Murai et al. (2018) were used:

$r = 0.06, \rho = 0.77, \lambda = 1.16 \ week^{-1}, \Gamma = 3.9 \ week^{-1}, \delta = 0.95$

5.3.1   Spatial simulations of p53 clone growth in back epidermis

Spatial simulations of p53 mutant cells growing within a wild type background achieved quantitative agreement with experimental observations (Figure 5.2). Without further additions or fitting of the model, p53 clone growth (Figure 5.2a) and tissue takeover (Figure 5.2b) match experimental observations, showing an initial exponential growth slowing over time.

Despite matching experimental data well, the simulations present a pathological behaviour. Double occupancies progressively accumulate in the basal layer over time. As a consequence, the overall tissue turnover slows and simulations can finish within the lifetime of the mouse when the simulated tissue becomes packed with double occupancies (Figure 5.2b,c,d), whereas the mouse tissue is found to have an increase in cell density of only $\sim 10\%$. Over time, the build up of double state cells leads to the increase of trapped differentiating cells which are unable to stratify. This reflects the imbalance in stem cell:differentiated cell ratio, eventually leading to drop in proliferation rate across the population (Figure 5.2e,f).

A typical simulation trajectory at different time points (0, 10, 20, 30 and 60 weeks), illustrated in Figure 5.3 provides a visual explanation on why the simulations fail to reproduce late time point dynamics.

The mutant population starts occupying space due to their property of producing descendants with decreased probability of leaving the tissue. The limited space availability impacts the growth of the wild type cells, which start forming double islands. In contrast to the wild type simulations, where double occupancies are transient events observed uniformly in the grid, in mutant simulations double cells are particularly notable in the mutated regions which become packed earlier. As a result, mutant cell clones expand up to a certain time point, when they become surrounded by double state wild type cells and their growth declines. The grid becomes heavily overcrowded, rendering the model unable to reach the lifespan of a mouse. Each grid space is occupied by two cells (double occupancies), either p53 mutant or wild type and as there is no space left for new division events, the simulation is terminated earlier than expected. These emergent properties imply that the tendency of the mutant population to produce persistent clones challenges the spatial SP model.
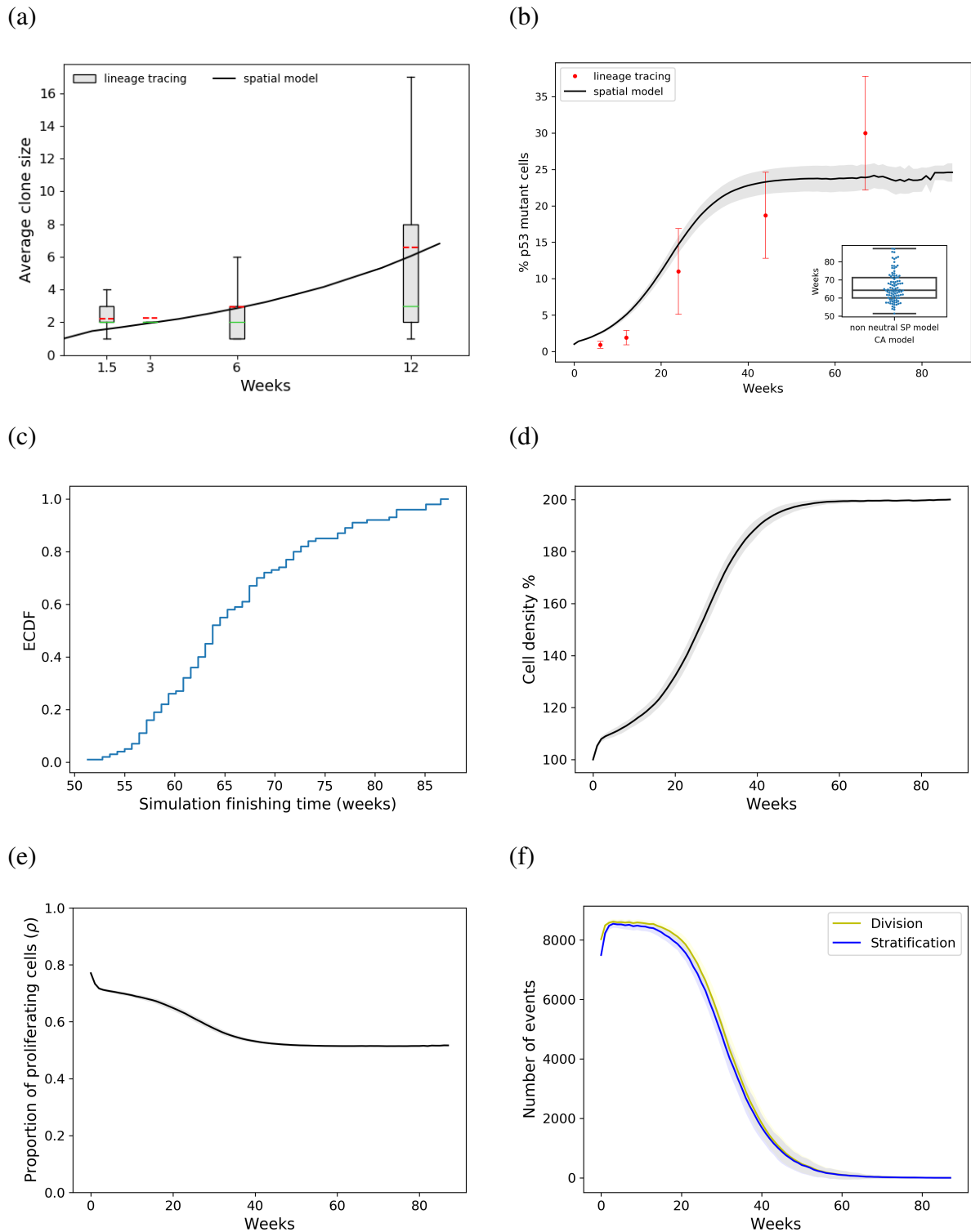
Fig. 5.2 Quantitative analysis of the spatial SP model including fate bias, simulating p53 mutant growth in the basal layer of back epidermis. Parameters: $r = 0.06$, $\rho = 0.77$, $\lambda = 1.16$/week, $\delta = 0.95$, from Murai et al. (2018). a) the simulated p53 average clone size matches observations, box plots correspond to experimental clone counts, dashed red lines: mean, solid green line: median, b) simulated and observed percentage of p53 mutant cells, inset: last simulated time point of each repetition (100 simulation repetitions were performed), c) Empirical cumulative density function (ECDF) of last simulated time point across 100 repetitions, d) cell density increases up to 100%, e) stem cell:differentiated cell ratio drops, f) the number proliferation and stratification events drops. a,b,d,e,f: Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.
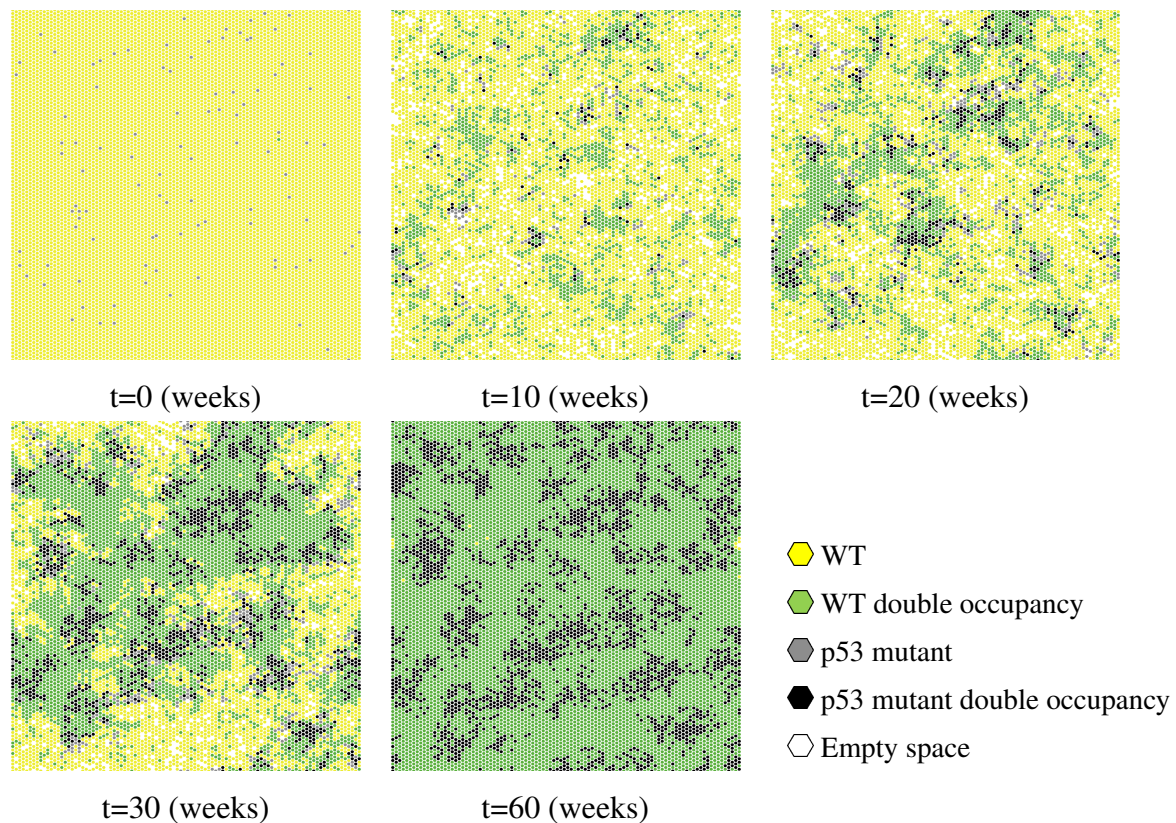
Fig. 5.3 Spatial non-neutral SP model time lapse. Typical simulation time lapse of the SP spatial model of p53 mutant growth. Cells are placed on a 2D 100x100 hexagonal lattice. Mutant behaviour fills the grid and slows growth across the basal layer, leading to 100% increase in cell density Parameters used: $r = 0.06$, $\rho = 0.77$, $\lambda = 1.16/\text{week}$, $\delta = 0.95$, from Murai et al. (2018).

A more detailed examination of the simulated p53 mutant population expansion (Figure 5.2b) revealed a slight overestimation of the mutant percentage compared to the experimental observations at the earliest time points. This reflects the relative uncertainty in calculating the initial induction level from the mutant exponential growth relative to the wild type situation. To further investigate the possibility of a lower mutant induction, I repeated the simulations of the spatial model testing a range of levels of p53 mutant induction, smaller than the reported 1% (Figure 5.4). Computational mutant induction level sweep suggested that a $\sim 0.04\%$ initial proportion of $p53^{*/wt}$ cells matches the experimental observations at the earliest time points. However, a striking output feature of the induction sweep analysis is that the model was consistently unable to fit last time point due to the unrealistic slow down in tissue turnover.
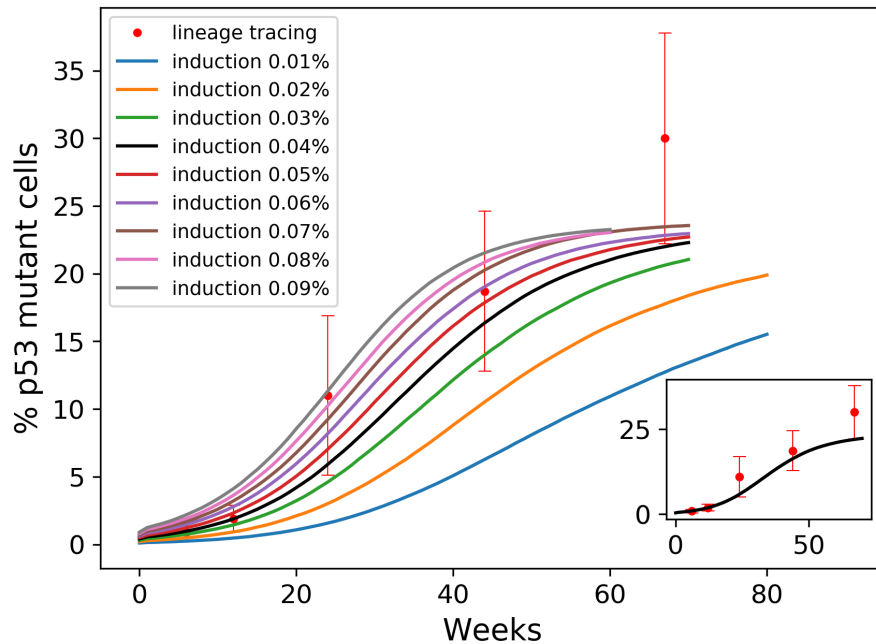
Fig. 5.4 Simulated percentage of the p53 mutant population under different initial proportions of mutant cells. An induction of $\sim 0.04\%$ improves the quantitative agreement between simulated and observed data at the two earliest time points. Despite that, the computational model consistently fails to fit last time point under all induction level regimes. The fit of a 0.04% p53 mutant induction (black curve) is also illustrated in the inset plot.

The above outlined observations clearly show that mutant behaviour fills the grid and slows growth across the whole tissue, leading to 100% increase in cell density, in contrast to experiments ($\sim 10\%$ increase) (Murai et al., 2018). The growth of p53 mutant clones approximates a logistic curve, but that without including limitations on mutation induced expansion the overall proliferation rate of the tissue drops due to space restrictions. As a result of this pathology, it is vital that we introduce feedbacks on stem cell fate choices that limit growth on the basis of crowding, despite the CA model's mutant takeover and average clone size being in agreement with the experiments.

5.3.2   p53 mutant logistic-like growth is confirmed by space-free model

To confirm that the tissue size was acting as a limiting factor and not the hexagonal grid, the model was re-implemented as a rule based model with a carrying capacity. Clonal data were generated using Bio-PEPA, a framework for modelling biochemical networks (Ciocchetta and Hillston, 2009). 10000 proliferating and differentiating cells, of which 1% corresponded to p53 mutants were initially considered (consistent to lattice dimensions) and population

dynamics was updated using fluid flow approximation. In fluid flow approximation discrete jumps are approximated by continuous flows between the states of the system and a set of ordinary differential equations (ODEs) is derived that approximate the average behaviour of the continuous time Markov Chains (CTMC) (Massink et al., 2010, Hillston, 2005). The resulted mutant cell populations were then compared with the ones derived from the lattice model (Figure 5.5).
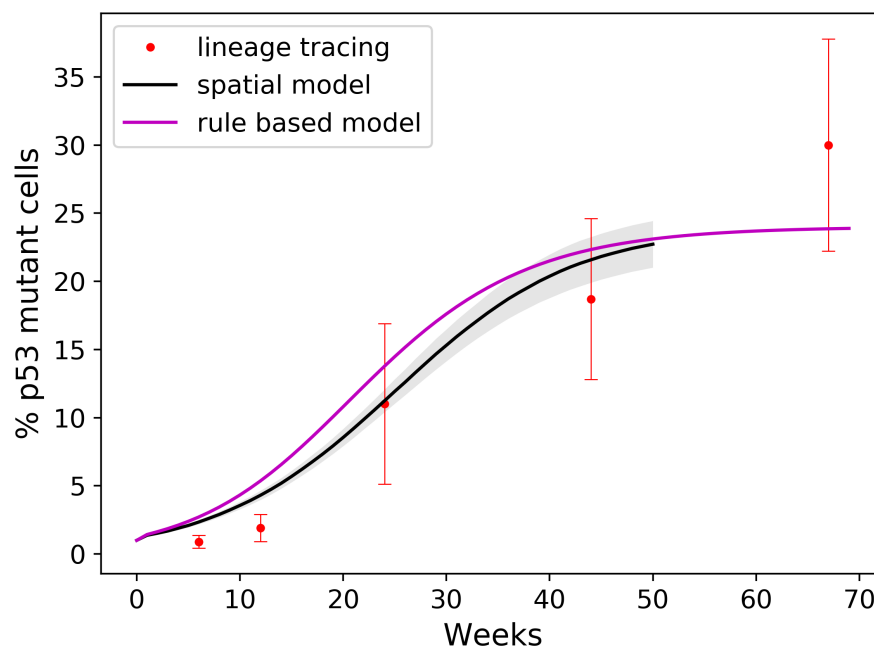
Fig. 5.5 Percentage of the p53 mutant population in spatial and rule based simulations with population limits. The two model types display similar behaviour. This suggests that logistic growth alone explains the experimentally observed behaviour of $p53^{*/wt}$ cells.

Interestingly, the two types of simulations display similar behaviour in mutant tissue takeover. As in lattice simulations, the proportion of mutant cells in the rule based model initially rises exponentially and then gradually slows down, reaching a plateau at late time points. A small shifting up of the mutant takeover curve produced by the rule based model indicates some kind of spatial influence. Local spatial constraints, imposed by the CA model, might affect the growth of individual clones, slightly slowing the takeover process. These simulations confirmed that whilst the presence of the grid slowed growth slightly, the growth curves were ultimately logistic-like. Therefore, the mutant growth behaviour on tissue scale appears to be constrained by tissue size. The reported phenotypic plasticity of p53 mutations in response to crowding matches logistic growth.

5.4 Introduction of spatial rules for feedback

Adding simple spatial rules for feedback could maintain tissue takeover and early clone growth whilst allowing tissue to turnover. In support of this, previous studies in epithelial tissues of different model organisms argue that local crowding and cell deformation induced by proliferation could trigger the delamination of nearby non-apoptotic cells (Eisenhoffer et al., 2012, Marinari et al., 2012, Miroshnikova et al., 2018). Furthermore, a recent study on mouse plantar epidermis suggests that division is stimulated by a stratification event in the local environment (Mesa et al., 2018). Additionally, theoretical studies have also stressed the importance of feedback mechanisms in regulating tissue growth (Shraiman, 2005, Vincent et al., 2013, Lee and Morishita, 2017). These spatial regulatory behavioural mechanisms between surrounding cells provide means to ensure tissue's homeostatic function by locally balancing stem cell density.

Considering the above reported response to crowding and the previously outlined model's pathological behaviour, the spatial model of p53 mutant growth was further refined to account for feedback effects controlling tissue growth.

5.4.1 Mechanical Feedbacks

Epithelial cells are constantly experiencing forces from their environment. The passive response of cells to mechanical forces generated within the tissue was initially considered. Two possible feedback rules were tested:

1. Diffusion: On the basis of the idea that spatial coordination mechanisms are seeking to minimize the local density imbalance within different regions of the tissue, cells were thought to be able to diffuse for a short distance. The induced stress from a division event could be released by the pushing of cells towards a nearby low density site.

   To include the rule of diffusion in the CA model, I made use of an extended neighbourhood. A dividing cell, instead of considering the six adjacent lattice sites for placing the new daughters, is able to exploit a vacant site in the expanded neighbourhood of twelve sites through the displacement of a cell from the immediate neighbour (Figure 5.6). If no empty space exists in the extended neighbourhood, the two new daughters are placed on the same space, creating a double occupancy.
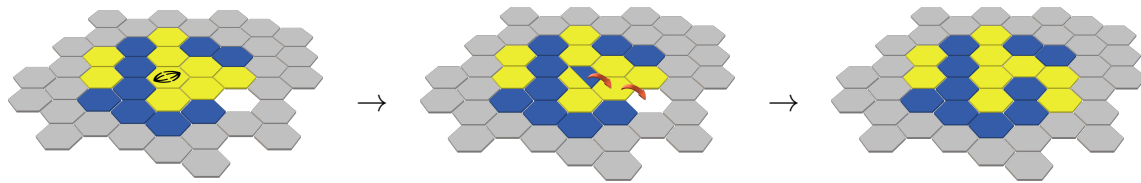
Fig. 5.6 Mechanical feedback of diffusion. A dividing cell seeks a vacant space in the extended neighbourhood of twelve lattice sites. If an empty space is available, an adjacent cell moves toward the vacant site, leaving space for the new daughters.

2. Crowding release: Motivated by the reports of cell extrusion events to compensate proliferation induced local stress (Eisenhoffer et al., 2012, Marinari et al., 2012, Miroshnikova et al., 2018, Frede et al., 2016), I introduced an analogous feedback response in my model. That is, overcrowding produced by double occupancies was released by allowing the immediate stratification of trapped differentiated cells.

   According to this feedback rule, every double occupancy consisting of at least one differentiating cell (AB, BA or BB double state cells) would lead to an instantaneous stratification event, allowing the release of stress induced by the double state cell (Figure 5.7).

(a)   (b)

(c)   (d)
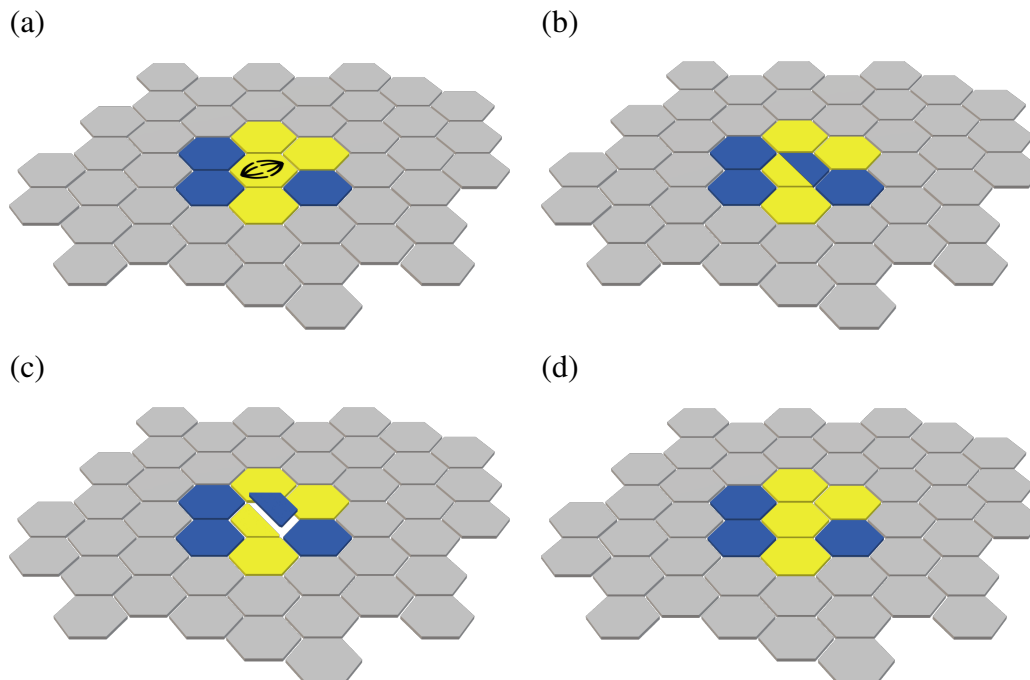
Fig. 5.7 Mechanical feedback of crowding release. A division event in a region without available adjacent empty space introduces a double occupancy (a). Since the division event lead to the production of a differentiating cell (AB or BB divisions) (b), one newly born differentiating daughter exits instantaneously the basal layer (c) to release the crowding (d).

Mechanical feedbacks successfully reproduce the homeostatic dynamics

To test whether the addition of such mechanical feedbacks has any impact on the expected SP homeostatic behaviour, each of the two new rules was initially applied to simulations of wild type epithelial basal cells. Figure 5.8 confirms that the spatial SP model with diffusion successfully reproduces the homeostatic hallmarks of cell population dynamics.

Similarly, the implementation of the spatial feedback rule with crowding release in wild type simulations revealed that the key features of epithelial homeostasis were successfully replicated (Figure 5.9). An observed characteristic property of the crowding release mechanism though is that the system maintains a constant basal layer density, a signature of homeostasis, following an initial $\sim 10\%$ decline (Figure 5.9e). However, as already mentioned, this does not alter the SP hallmarks.

Importantly, the inclusion of either the diffusion or the crowding release feedback in wild type simulations appears to reduce the cell density fluctuations observed in the initial spatial SP model, as the models with feedbacks produce a less broad distribution in basal cell density compared to that of the model without feedbacks (Figures 5.8f, 5.9f). Furthermore, the cell density distribution in the feedback models qualitatively appears to resemble more with that of microscopy images (Figures 5.8f, 5.9f).
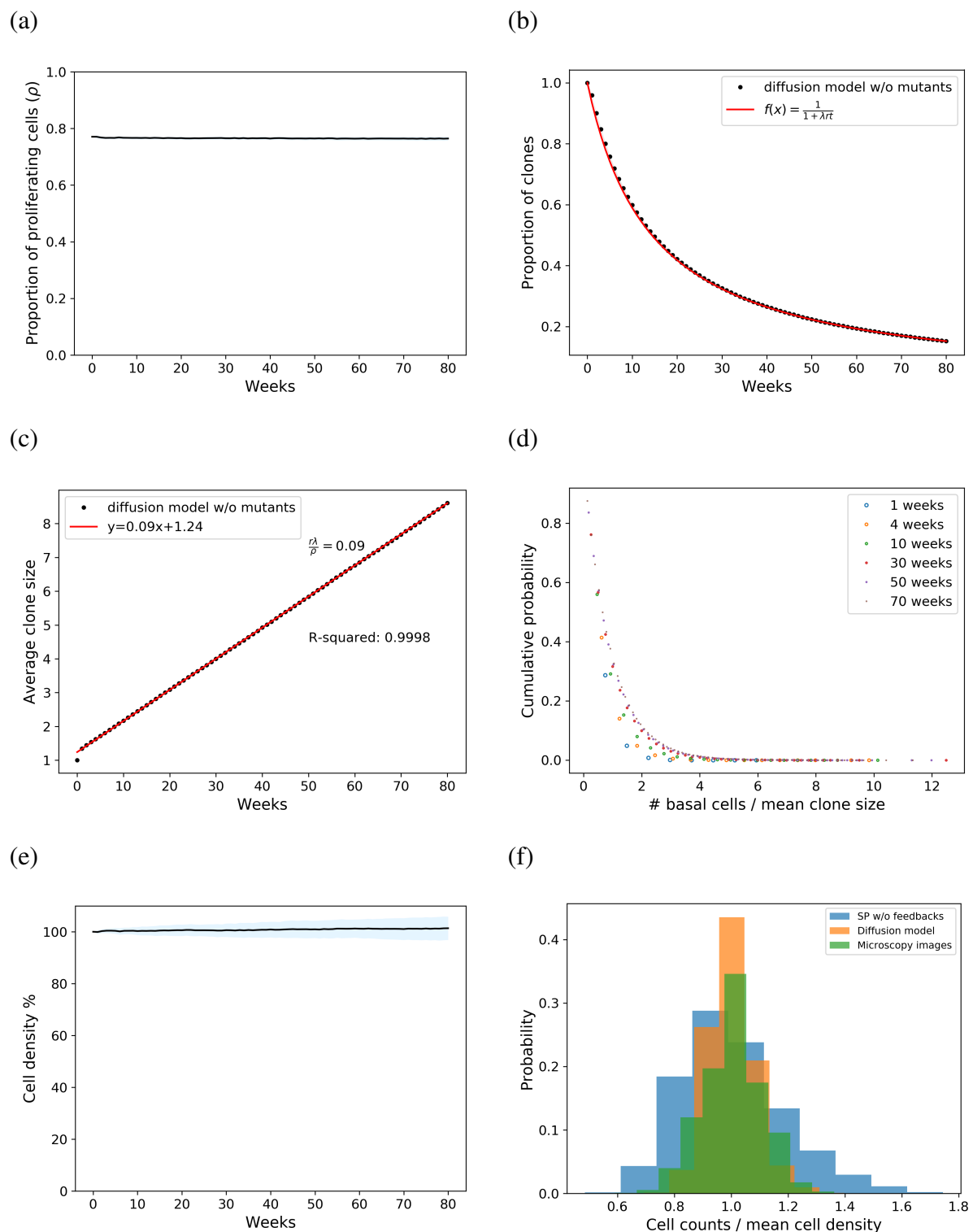
Fig. 5.8 Quantitative analysis of the spatial fate bias SP model with diffusion, not including mutant cells, in the basal layer of back epidermis. Parameters: $r = 0.06$, $\rho = 0.77$, $\lambda = 1.16/\text{week}$, from Murai et al. (2018). Key homeostatic properties replicated: a) proportion of progenitor cells remains constant, b) the number of surviving clones decreases over time following $\frac{1}{1+\lambda rt}$, c) the average clone size increases over time with slope $\tau = \frac{r\lambda}{\rho}$, d) clone size distribution scales with time, e) overall basal layer density remains constant, f) cell density distributions in confocal microscopy images of mouse oesophagus, in spatial SP model with diffusion and in spatial SP model without cell density dependent feedbacks (t = 52 weeks). a,b,c,d,e: Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

(a)

(b)

(c)

(d)

(e)

(f)
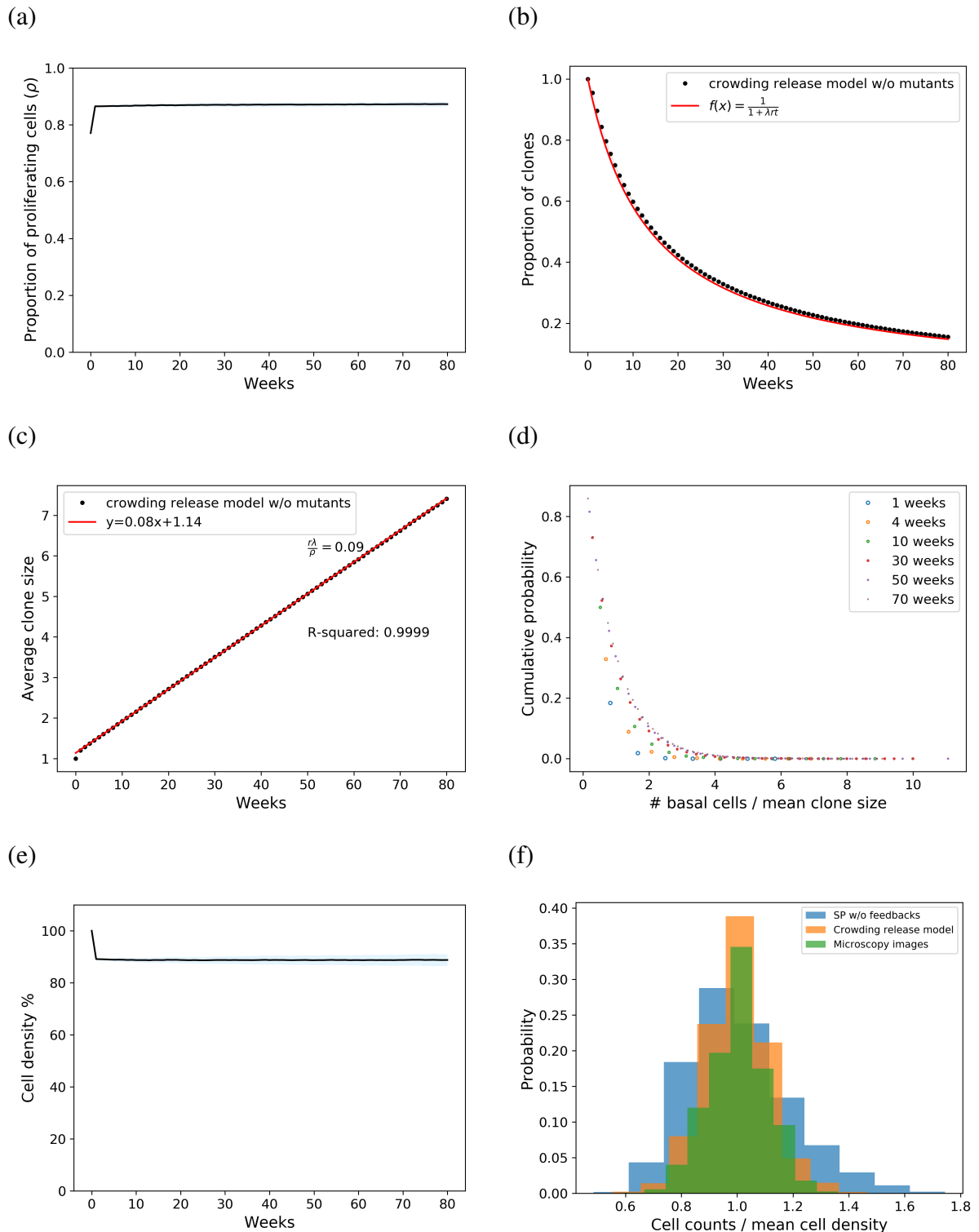
Fig. 5.9 Quantitative analysis of the spatial fate bias SP model with crowding release, not including mutant cells, in the basal layer of back epidermis. Parameters: $r = 0.06$, $\rho = 0.77$, $\lambda = 1.16$/week, $\delta = 0.95$, from Murai et al. (2018). Key homeostatic properties replicated: a) proportion of progenitor cells remains constant, b) the number of surviving clones decreases over time following $\frac{1}{1+\lambda rt}$, c) the average clone size increases over time with slope $\tau \sim \frac{r\lambda}{\rho}$, d) clone size distribution scales with time, e) overall basal layer density drops during the first week but afterwards it remains constant, f) cell density distributions in confocal microscopy images of mouse oesophagus, in spatial SP model with crowding release and in spatial SP model without cell density dependent feedbacks (t = 52 weeks). a,b,c,d,e: Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

p53 mutant simulations including the diffusion feedback

Computational simulations of the spatial SP model with p53 mutant cells including the feedback of diffusion were performed. As demonstrated in Figures 5.10, 5.11, considering a wider neighbourhood for placing the newly born cells does not achieve any improvement in reproducing the experimental observations and releasing the crowding frustration. On the contrary, the diffusion feedback mechanism facilitates mutant expansion as p53 cells are less restricted in their spreading. A visual inspection of the grid (Figure 5.10) highlights this issue, illustrating a sharp increase in the incidence of double occupancies between weeks 30 and 40. Analysis of p53 mutant growth dynamics revealed that, whilst the average clone size matched the data (Figure 5.11a), the mutant population increased rapidly (Figure 5.11b), resulting in the tissue being overcrowded faster (Figure 5.11c,d). The emergence of an increased density system forces the simulations to terminate earlier than expected, with the majority of the simulations being unable to reach the average mouse lifespan of ~ 80 weeks (Figure 5.11e).
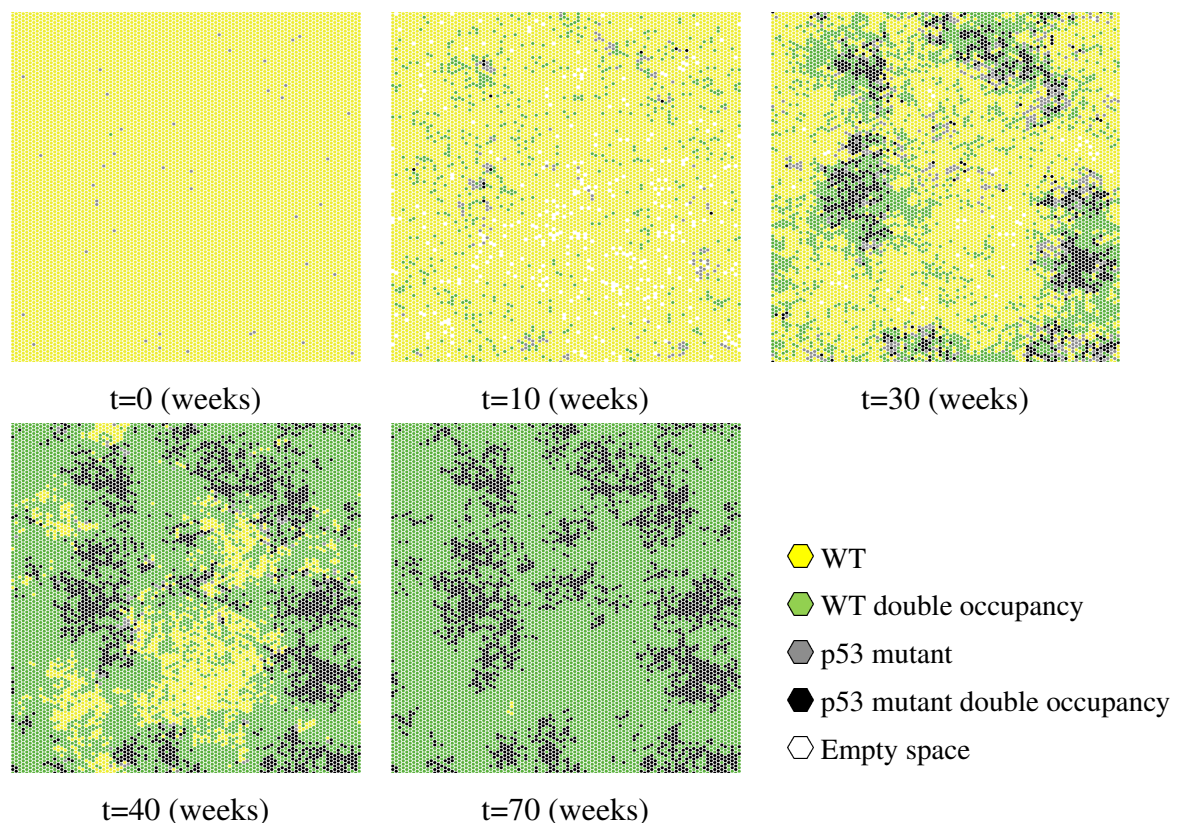
Fig. 5.10 Spatial non-neutral SP model with diffusion time lapse. Typical simulation time lapse of the SP spatial model of p53 mutant growth with diffusion. Cells are placed on a 2D 100x100 hexagonal lattice. Mutant population expands rapidly, promoting an acute increment of double occupancies. Parameters used: $r = 0.06$, $\rho = 0.77$, $\lambda = 1.16$/week, $\delta = 0.95$, from Murai et al. (2018).
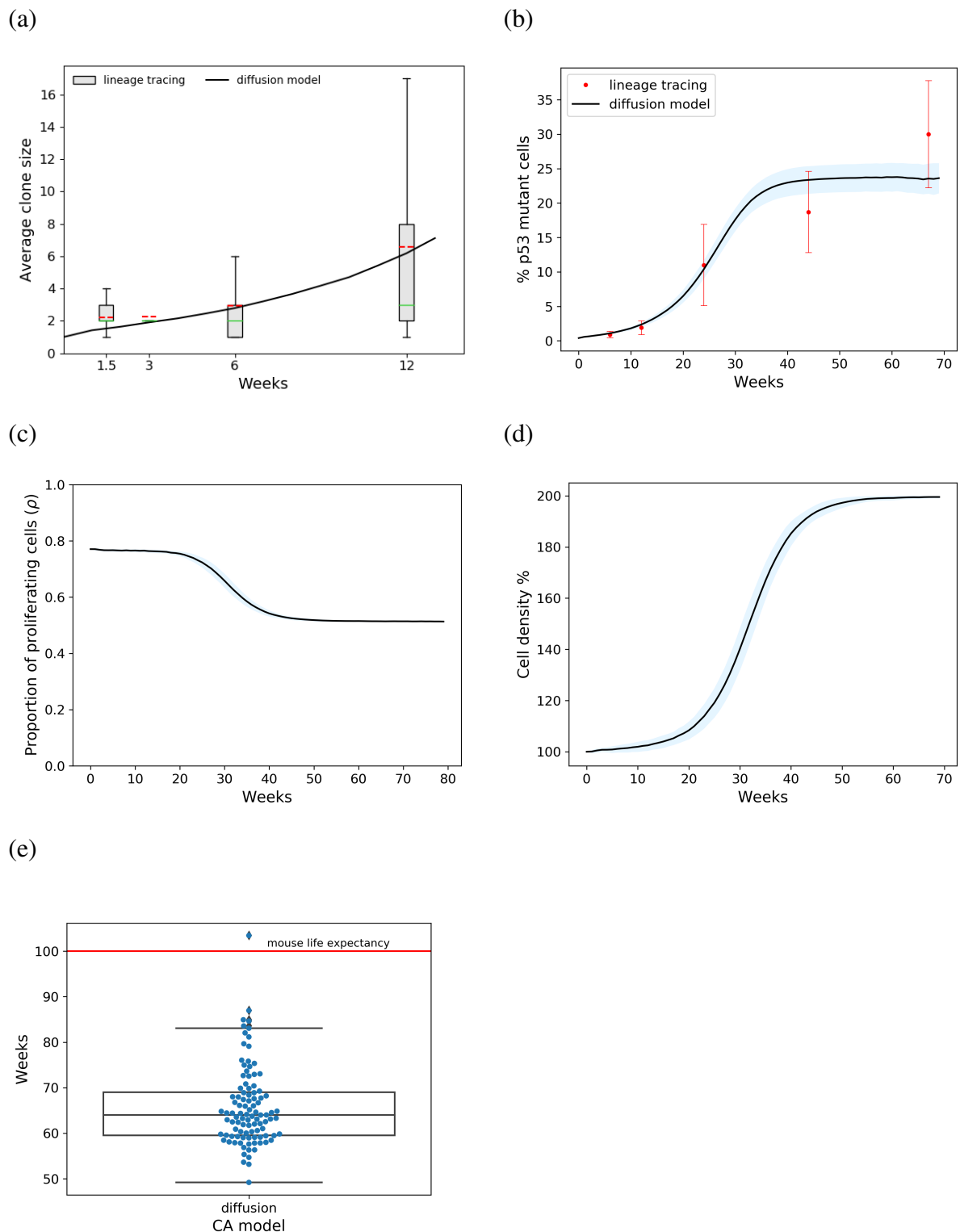
Fig. 5.11 Quantitative analysis of the spatial fate bias SP model with diffusion, simulating p53 mutant growth in the basal layer of back epidermis. Parameters: $r = 0.06$, $\rho = 0.77$, $\lambda = 1.16/\text{week}$, $\delta = 0.95$, from Murai et al. (2018). a) the simulated p53 average clone size matches observations, box plots correspond to experimental clone counts, dashed red lines: mean, solid green line: median, b) simulated and observed percentage of p53 mutant cells, c) stem cell:differentiated cell ratio drops, d) cell density sharply increases up to 100%, e) last simulated time point of each repetition (100 simulation repetitions were performed). a,b,c,d: Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

p53 mutant simulations including the crowding release feedback

The alternative feedback mechanism of crowding release was also explored. The inclusion of instantaneous stratification of differentiating cells in regions of overcrowding (double occupancies) largely succeeds in releasing the stress imposed by mutant growth. This is confirmed by a smaller rise in cell density (Figures 5.12a), which allows a substantial increase in the duration of the computational simulations beyond a mouse lifetime (Figures 5.12b). Nevertheless, this spatial rule causes an initial $\sim 10\%$ drop in cell density (Figure 5.12a) and a reduction in the number of differentiating cells, demonstrated by an increase in ρ value (Figure 5.12c). The aforementioned emergent property, also observed in wild type simulations (Figures 5.9a,e), might create more space for the $p53^{*/wt}$ cells to grow, leading to an enhanced takeover at later time points (Figure 5.12d), which in contrast to the average clone size (Figure 5.12e) does not match the experimental observations. Figure 5.13 illustrates a typical simulation trajectory where the crowding release mechanism was applied.

Taken together, the previously analysed feedback mechanisms could not successfully reproduce the exact behaviour of the system. The crowding release model tends to overestimate the mutant takeover whereas the diffusion model leads to a rapid expansion of mutant clones resulting in the simulated tissue becoming overcrowded very quickly. These findings indicate that the p53 mutant takeover dynamics is insufficiently explained solely by the pure effect of physical forces. The scenario where cells communicate with their neighbours via molecular signalling and make fate choices in response to local stress should be considered.

(a)

(b)

(c)

(d)

(e)
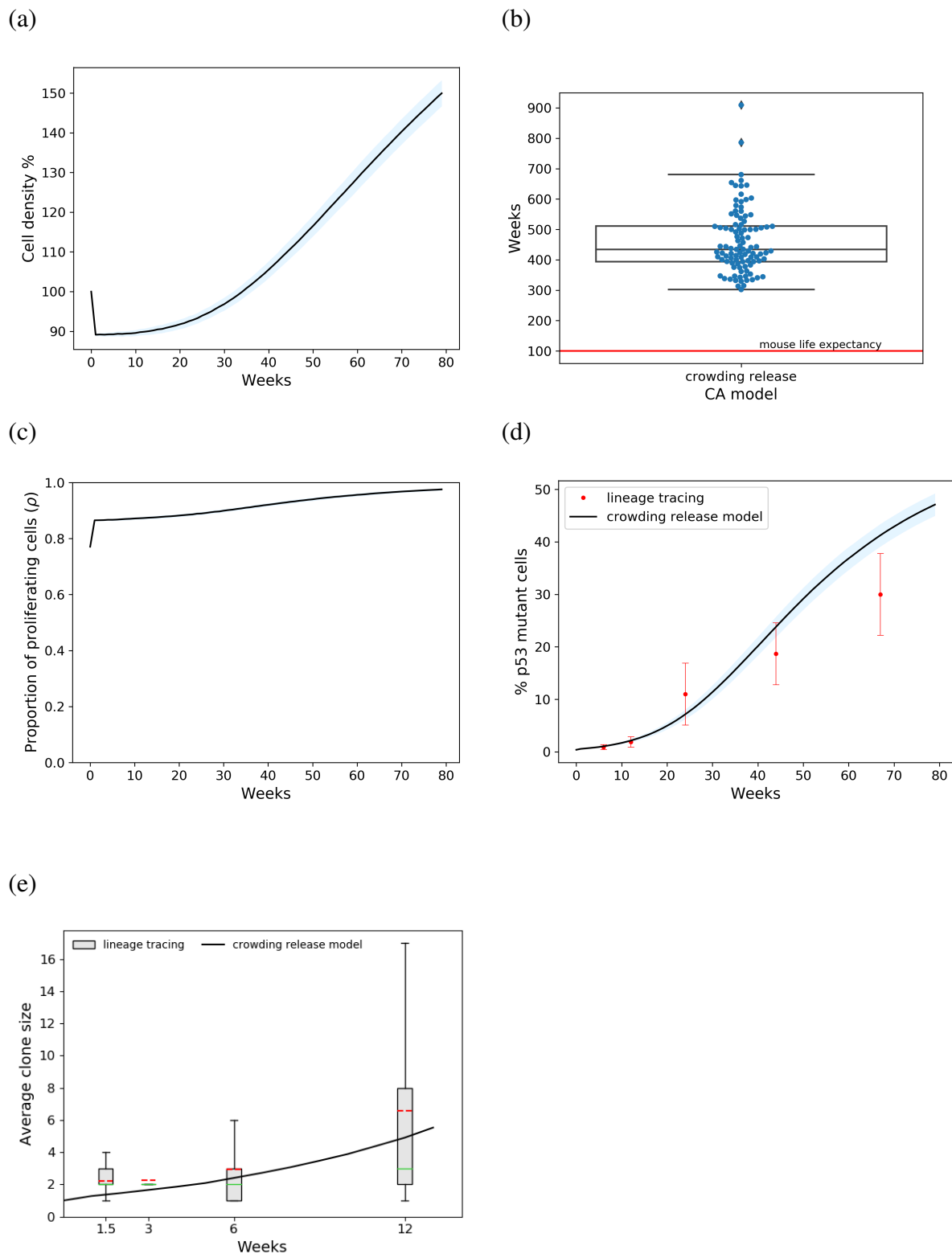
Fig. 5.12 Quantitative analysis of the spatial SP fate bias model with crowding release, simulating p53 mutant growth in the basal layer of back epidermis. Parameters: $r = 0.06$, $\rho = 0.77$, $\lambda = 1.16$/week, $\delta = 0.95$, from Murai et al. (2018). a) Cell density increases up to 50% following an initial 10% drop, b) last simulated time point of each repetition (100 simulation repetitions were performed), c) stem cell:differentiated cell ratio increments, d) an overestimation in simulated tissue takeover compared to experimental values is observed, e) the simulated p53 average clone size matches observations, box plots correspond to experimental clone counts, dashed red lines: mean, solid green line: median. a,b,c,d: Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.
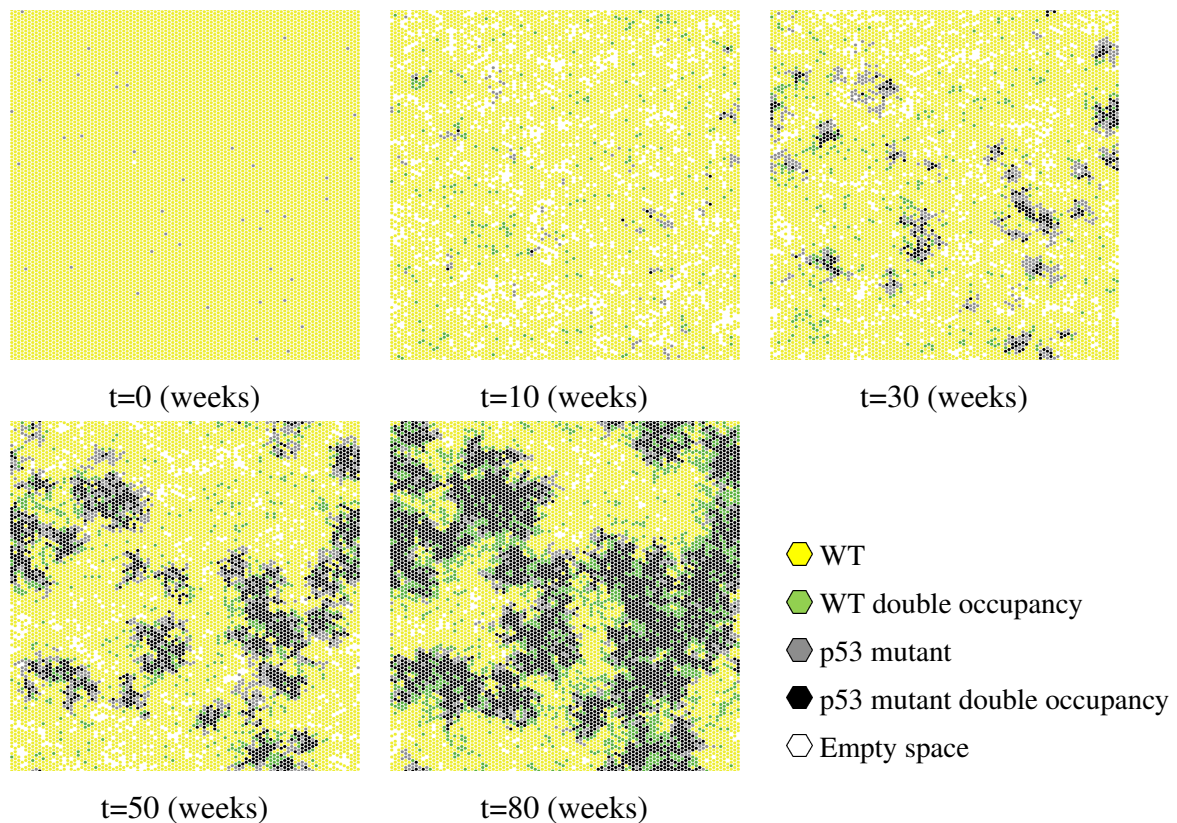
Fig. 5.13 Crowding release time lapse. Typical simulation time lapse of the SP spatial model of p53 mutant growth with crowding release. Cells are placed on a 2D 100x100 hexagonal lattice. The proportion of double occupancies is substantially reduced, releasing the stress imposed by mutant growth.

5.4.2   Cell - cell communication feedbacks

In seeking to determine a more suitable spatial rule to simulate p53 mutant behaviour, achieving both fit to experimental observations and tissue survival, feedbacks based on cell-cell communication signalling were introduced. The new feedback rules were based on the hypothesis that cells sense the density of their immediate neighbourhood and depending on how crowded their local environment is, a bias towards the production of either proliferating or differentiating progeny is decided. Such a feedback mechanism would appear consistent with previous studies reporting density dependent spatial regulation in epithelial tissues (see 5.4).

Two classes of feedback mechanisms were tested, termed "upstream" and "downstream" to reflect which cells are affected by the feedback and their relationship to the mutations under study (Figure 5.14). In the upstream feedback, all cells respond to crowding events by inducing an additional imbalance in cell fate to promote differentiation, and cells with

mutations sum their imbalances. This arises as the genes changed in response to crowding events are either independent or upstream of the mutated genes. In the downstream feedback, only mutated cells respond to crowding events as the gene target being altered by crowding is considered to be downstream of the mutant gene. In each case, a threshold for activation was set based on the total sum of immediate neighbours surrounding a cell, treating double occupancies as two cells.
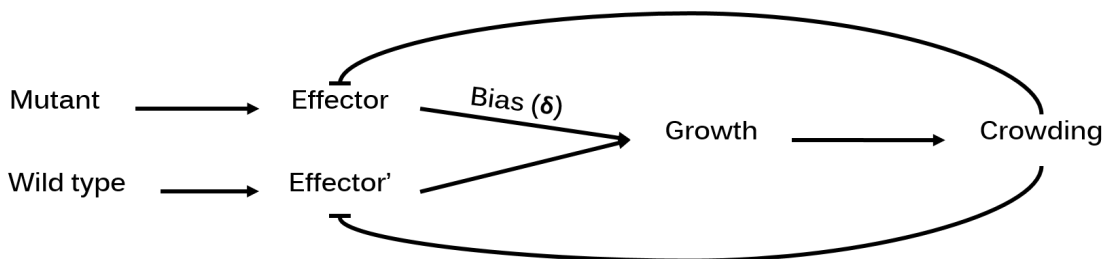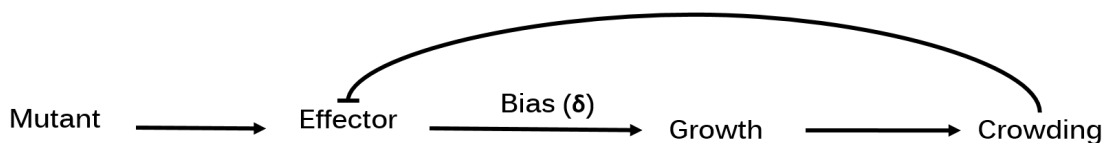
Fig. 5.14 **Feedbacks based on cell-cell communication.** a) The "upstream" feedback hypothesis assumes that genes changed in response to crowding are independent of the mutated genes, b) the "downstream" feedback hypothesis suggests that the genes altered by crowding are downstream of the mutated genes and therefore, only mutant cells respond to crowding events.

The upstream feedback enhances tissue takeover

Initially, the upstream feedback mechanism was tested. An additional fate bias parameter, δ', was introduced to the fate imbalance spatial SP model. The value of δ' depends on the local cell density, i.e. number of neighbouring cells. A neighbourhood consisting of more than a defined number of cells would be considered as "crowded" whereas fewer neighbouring cells than the defined crowding cut-off would indicate an underpopulated region ("empty"). In the former case, δ' is negative rendering the fate of dividing cells tilted towards symmetric differentiation to release crowding. In the latter case, a positive δ' is chosen, favouring symmetric division to fill the empty sites. When the neighbourhood is neither "crowded" nor "empty", it indicates that the local cell density is homeostatic and therefore $\delta' = 0$ (Figure 5.15).

(a)

(b)

(c)
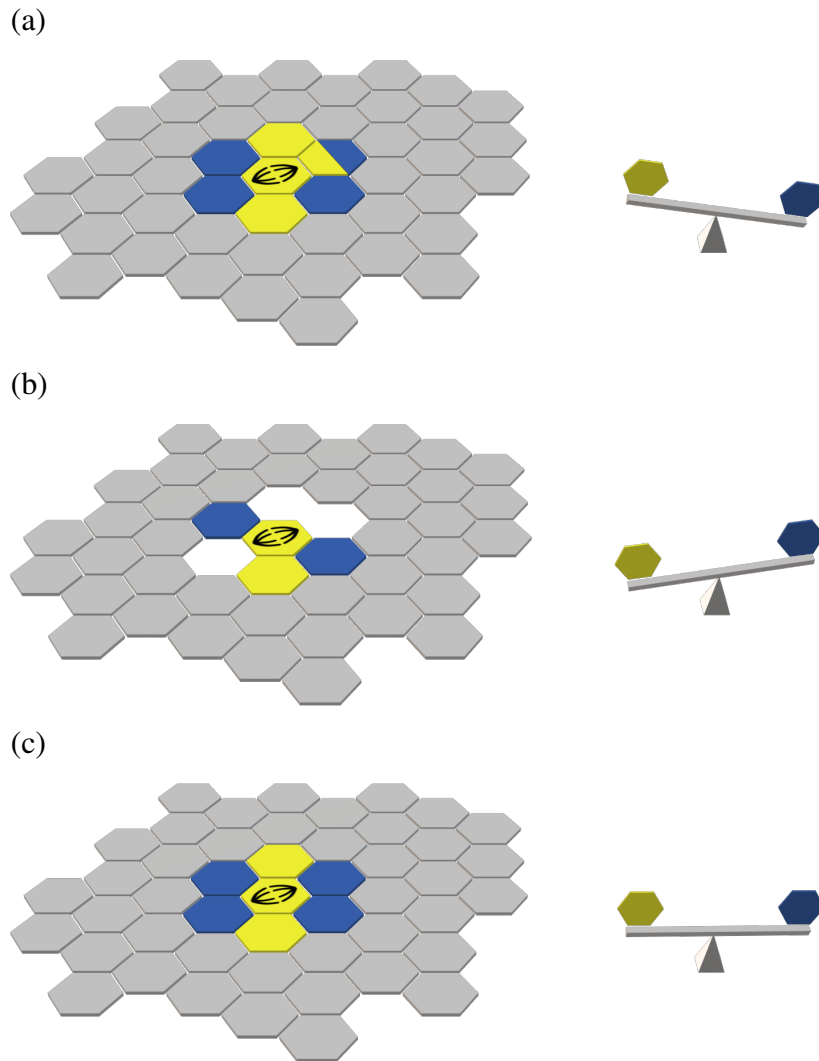
Fig. 5.15 Upstream feedback. The fate decision process in every simulated cell is affected by the number of neighbouring cells. a) When a dividing cell is surrounded by a crowded environment, it is prone to produce differentiating progeny, promoting local stress release, b) when the neighbourhood of a dividing cell is underpopulated, a preference towards the production of proliferating progeny is decided, to fill the empty space, c) when the neighbourhood of a dividing cell is homeostatic, i.e. neither crowded nor empty, division probabilities are balanced. Dividing cells are displayed in yellow, differentiating cells are displayed in blue.

$$A \varnothing \xrightarrow{\lambda} \begin{cases} AA & r(1 + \delta + \delta') \\ AB & \frac{1}{2} - r \\ BA & \frac{1}{2} - r \\ BB & r(1 - \delta - \delta') \end{cases}$$

$$A X \xrightarrow{\lambda} \begin{cases} D_{AA} & r(1 + \delta + \delta') \\ D_{AB} & \frac{1}{2} - r \\ D_{BA} & \frac{1}{2} - r \\ D_{BB} & r(1 - \delta - \delta') \end{cases} \tag{5.3}$$

$$\delta' \to \begin{cases} + & n < \text{cell density cut-off} \\ 0 & n = \text{cell density cut-off} \\ - & n > \text{cell density cut-off} \end{cases}$$

$$B \xrightarrow{\Gamma} \varnothing,$$

where $\delta' \in [0, 1]$. As each cell on the grid has six neighbours in normal conditions, i.e. no overcrowding, no gaps, the local cell density crowding cut-off was initially set to six cells. Furthermore, as wild type cells do not have a δ, they would solely respond to the cell density bias (δ'), whereas in the case of p53 mutant cells the cell density bias (δ') and their innate fate bias (δ) would be counterbalanced (5.3).

Initially, simulations not introducing mutant cells were performed and confirmed that the upstream feedback regime is able to replicate the SP signatures on cell population dynamics (Figure 5.16). Notably, the inclusion of this feedback mechanism reduced the local cell density heterogeneity previously observed in the initial SP spatial wild type simulations by producing a less broad basal cell density distribution (Figure 5.16f). Subsequently, the new feedback model was applied to mutant tissue simulations and was tested on a set of δ' parameter values (Figure 5.17).

(a)

(b)

(c)

(d)

(e)

(f)
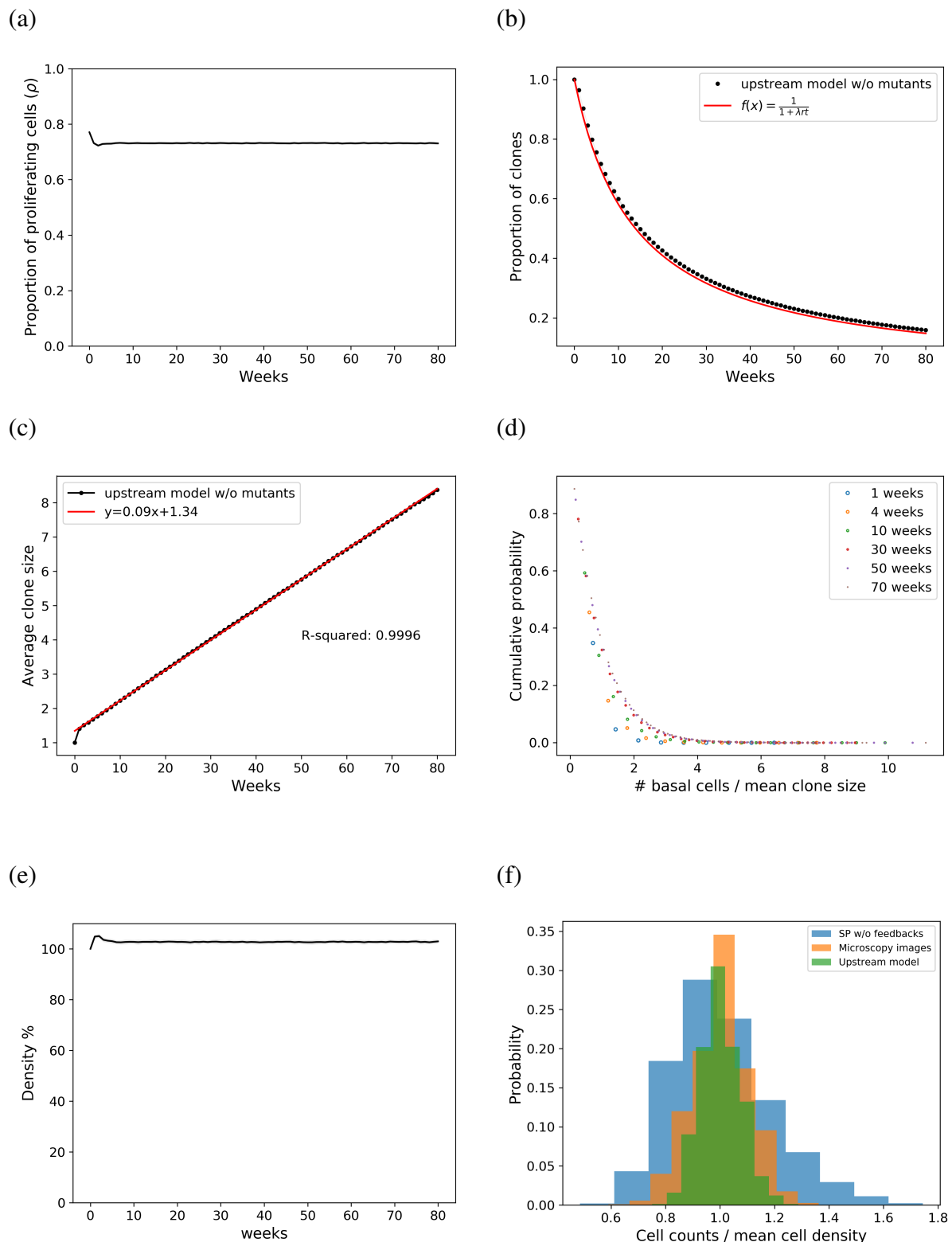
Fig. 5.16 Quantitative analysis of the spatial fate bias SP model with upstream feedbacks, not including mutant cells, in the basal layer of back epidermis. Parameters: $r = 0.06$, $\rho = 0.77$, $\lambda = 1.16$/week, from Murai et al. (2018). Key homeostatic properties replicated: a) proportion of progenitor cells remains largely constant, b) the number of surviving clones decreases over time following $\frac{1}{1+\lambda rt}$, c) the average clone size increases over time with slope $\tau \sim \frac{r\lambda}{\rho}$, d) clone size distribution scales with time, e) overall basal layer density remains constant, f) cell density distributions in confocal microscopy images of mouse oesophagus, in spatial upstream feedback SP model and in spatial SP model without cell density dependent feedbacks (t = 52 weeks). a,b,c,d,e: Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

(a)

(b)

(c)
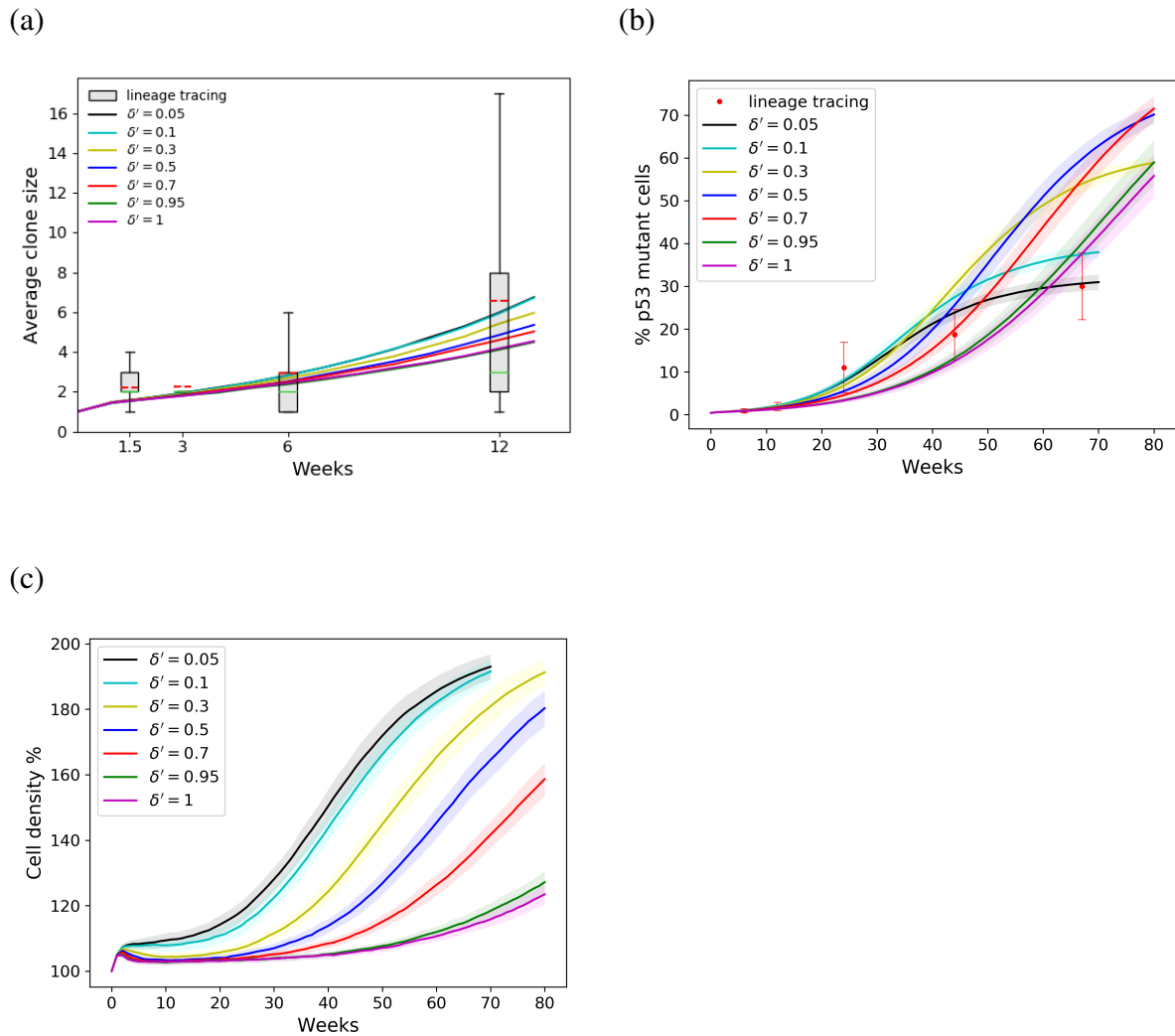
Fig. 5.17 Quantitative analysis of the spatial fate bias SP model with upstream feedbacks, simulating p53 mutant growth in the basal layer of back epidermis. Parameters: $r = 0.06$, $\rho = 0.77$, $\lambda = 1.16/\text{week}$, $\delta = 0.95$, from Murai et al. (2018). Simulated results across different δ' levels: a) simulated p53 average clone size matches observations, box plots correspond to experimental clone counts, dashed red lines: mean, solid green line: median, b) simulated and observed percentage of p53 mutant cells, c) simulated overall cell density. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

As shown in Figure 5.17, the upstream regime cannot achieve both fit to the experimental observations (mutant takeover and tissue cell density) and stress release. Smaller δ' values show much closer agreement to the experimental values of average clone size and mutant tissue takeover (Figures 5.17a,b), but they are unable to achieve crowding release (Figure 5.17c). On the contrary, when applying higher δ' values, a more realistic tissue overall cell density is accomplished (Figure 5.17c), but a much higher proportion of mutant cells is observed (Figure 5.17b). This indicates that a stronger density feedback, applied to

all cells in the tissue, makes competition last longer as it maintains the fitness advantage of the mutant population. A much more fluid tissue is generated which leads to an increase in mutant takeover beyond the values observed in the experiments. On the other hand, a weaker density feedback has almost the same effect as the no feedback simulations would have, that is, closer match to mutant takeover but growth decline due to overcrowding. Taken together, the inclusion of a spatial feedback applied upstream to mutant cells enhances tissue takeover and therefore is inappropriate to describe the dynamics of p53 mutants. A different approach of applying cell-cell communication feedbacks should then be adopted.

Downstream feedbacks appropriately describe mutant dynamics

As an alternative hypothesis, the downstream feedback, where the response to a crowded neighbourhood is activated downstream to p53 mutations mechanisms was tested. This implies that whilst $p53^{*/wt}$ cells initially grow unrestricted and outcompete wild type cells, they eventually lose their proliferative advantage due to space restrictions. The downstream feedback hypothesis would be more consistent with the observed phenotypic plasticity of p53 mutant population in response to crowding (Murai et al., 2018).

To implement the new feedback mechanism, no additional cell density fate bias parameter (δ') was used but p53 mutants were set to lose their innate bias towards the production of proliferating progeny, (δ), as a reaction to stress induced in their local neighbourhood. Thus, the downstream feedback model can be described by (5.2), page 115 of a spatial model with fate bias. The value of δ parameter is turned off when the local cell density (i.e. number of cells) of a dividing $p53^{*/wt}$ cell's neighbourhood consists of more than six cells, switching the mutant behaviour to balanced mode ($\delta = 0$).

(a)

(b)

(c)

(d)
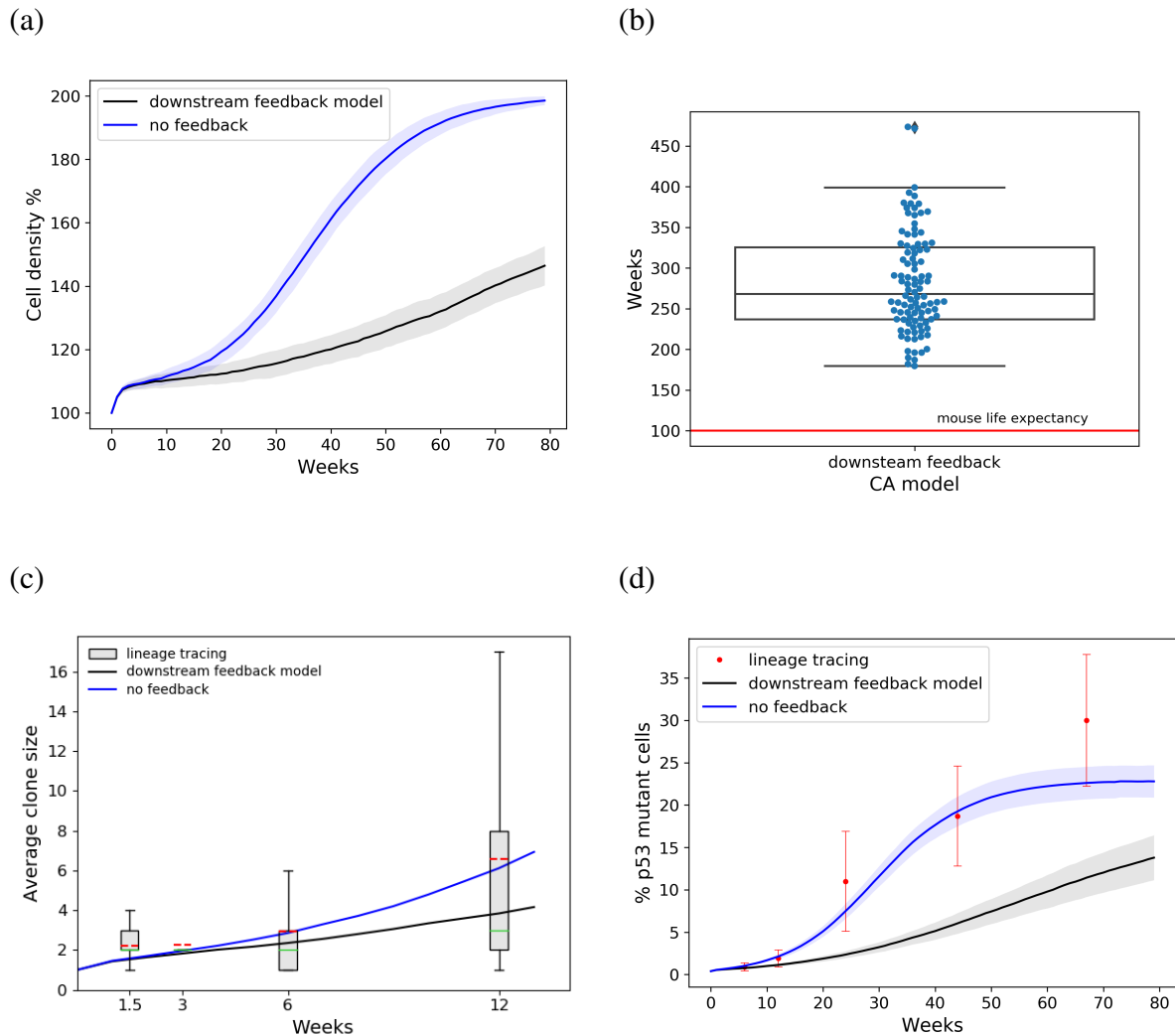
Fig. 5.18 Quantitative analysis of the spatial fate bias SP model with downstream feedbacks, simulating p53 mutant growth in the basal layer of back epidermis. Parameters: $r = 0.06$, $\rho = 0.77$, $\lambda = 1.16$/week, $\delta = 0.95$, from Murai et al. (2018). Black curve: fate bias SP model with downstream feedbacks, blue curve: fate bias SP model without feedbacks: a) the increase in the simulated overall basal cell density at 67 weeks (end of experiment) is closer to the experimentally measured 10% increase, b) last simulated time point of each repetition (100 simulation repetitions were performed), c) the simulated p53 average clone size, box plots correspond to experimental clone counts, dashed red lines: mean, solid green line: median, d) model underestimates the percentage of p53 mutant cells. a,c,d: Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

To explore the spatial system of interacting mutant and wild type cells under the downstream feedback scenario, computational simulations were performed. The spatial model of p53 mutant growth with downstream feedback response produced a substantially less crowded basal layer compared to the mutant model with no feedbacks. The cell density increase at 67 weeks was closer to but did not match the experimentally measured 10%

at the same time point (Figure 5.18a). This enabled model simulations to finish at time points beyond a mouse's lifetime (Figure 5.18b). However, an underestimation of mutant average clone size and tissue takeover was observed. Whilst the simulated $p53^{*/wt}$ average clone size was lower than the experimental mean but within the interquartile range (IQR) of clone size distribution (Figure 5.18c), the $p53^{*/wt}$ percentage was not in agreement with the experiments (Figure 5.18d).

Two different interpretations of the reduced takeover may exist. One possible reason would be that mutant cells might respond to a different crowding level. The emergence of stress experiencing areas in the neighbourhood of mutant cells at early time points might also explain the limited takeover. Mutant p53 cells tend to be surrounded by regions of double occupancies (Figure 5.19), resulting in their restrained expansion as they lose their growth advantage.
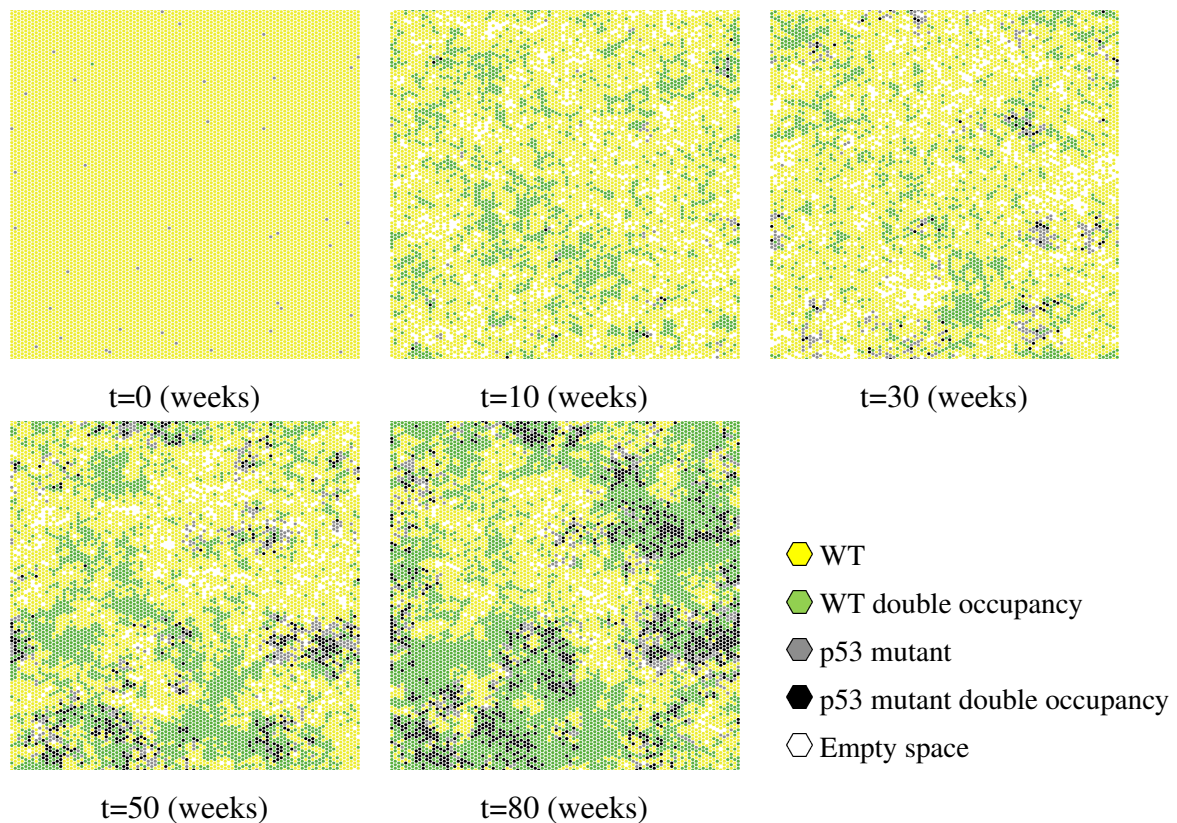
Fig. 5.19 Downstream feedback model time lapse. Typical simulation time lapse of the SP spatial model of p53 mutant growth with downstream feedbacks. Cells are placed on a 2D 100x100 hexagonal lattice. Mutant p53 cells tend to be surrounded by regions of double wild type cells, resulting in their restrained expansion as they lose their growth advantage.

Driven by these assumptions, the rules of the downstream feedback model were slightly modified to explore whether the issue in the simulated mutant takeover could be resolved.

The speculation that the mutant cells activate their response to stress at a higher than the initially considered 6-cell density was investigated. To include this in the model, simulations were repeated using less strict neighbourhood crowding cut-offs. I tried three different density cut-off values, 7, 8 and 9 cells.

(a)

(b)

(c)

(d)
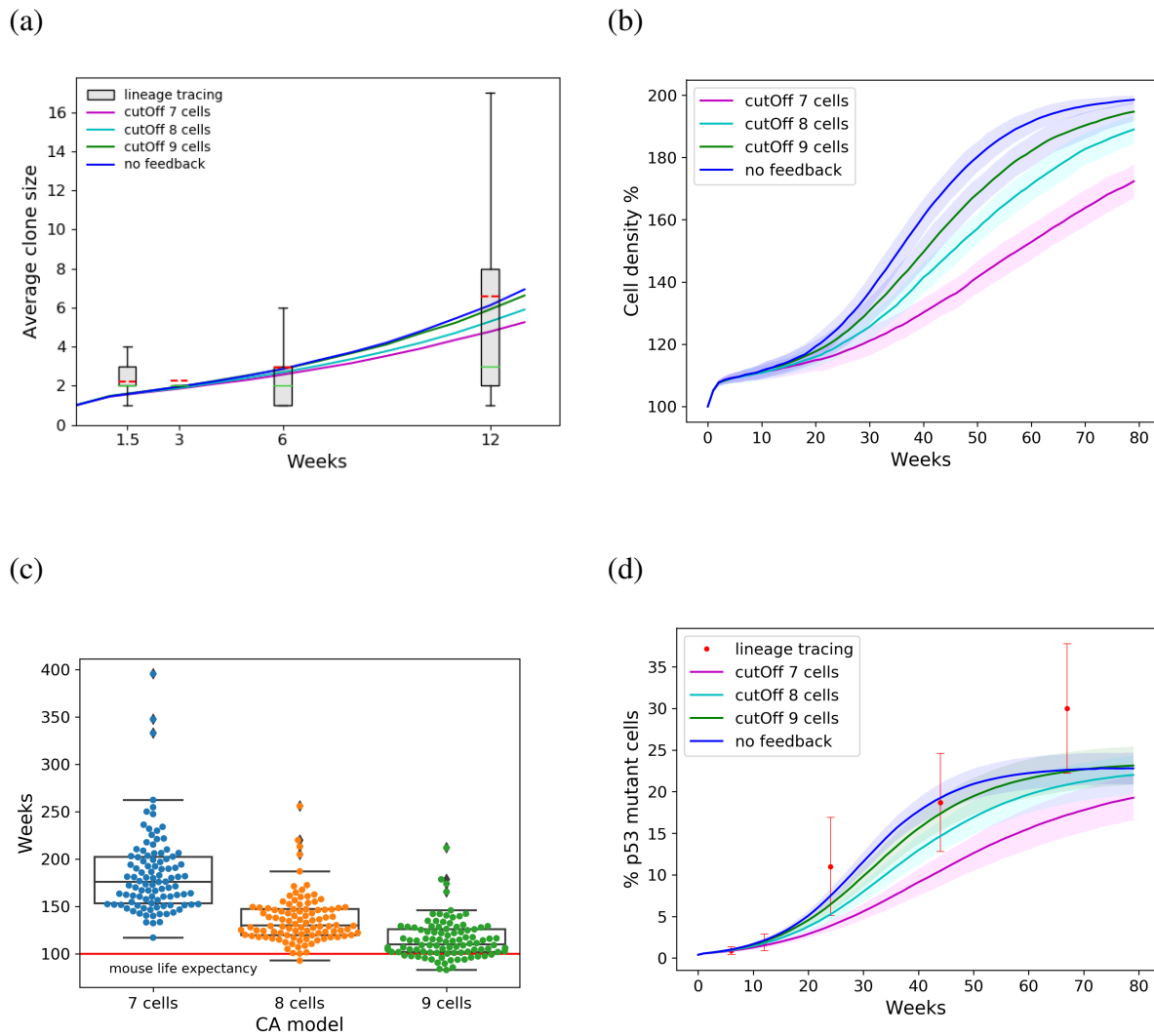
Fig. 5.20 Quantitative analysis of the spatial fate bias SP model with downstream feedbacks using different crowding cut-offs, simulating p53 mutant growth in the basal layer of back epidermis. Parameters: $r = 0.06$, $\rho = 0.77$, $\lambda = 1.16$/week, $\delta = 0.95$, from Murai et al. (2018). a) the simulated p53 average clone size broadly fits experimental observations, box plots correspond to experimental clone counts, dashed red lines: mean, solid green line: median, b) simulated overall basal cell density increases at higher thresholds, c) last simulated time point of each repetition (100 simulation repetitions were performed), d) simulated and observed percentage of p53 mutant cells. Lower thresholds underestimate data. a,b,d: Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

Clonal analysis of the downstream feedback model with varied cell density cut-offs showed that none of them is able to give both agreement with experimental observations and crowding release. All tested crowding thresholds produce acceptable fits in terms of the average clone size (Figure 5.20a). Nevertheless, a more relaxed crowding cut-off value leads to an increased basal layer cell density (Figure 5.20b). Despite the 7 cell cut-off simulations being able to maintain tissue turnover for a longer time period (Figure 5.20c), they appear to underestimate the mutant population proportion (Figure 5.20d). This indicates that decreasing the mutant cell population sensitivity to crowding by using a higher neighbourhood cell density threshold does not appear to be compatible to the experimental observations.

As already pointed out, another possible limiting factor of mutant expansion is the crowded areas (i.e. accumulation of double occupancies) at their clone boundaries. To overcome this, I repeated the simulations including the crowding release rules, where B type cells in double occupancies are allowed to stratify instantaneously, as described in Section 5.4.1, page 123. A typical simulation trajectory of the model with the two feedbacks combined is able to successfully maintain tissue turnover (Figure 5.21).
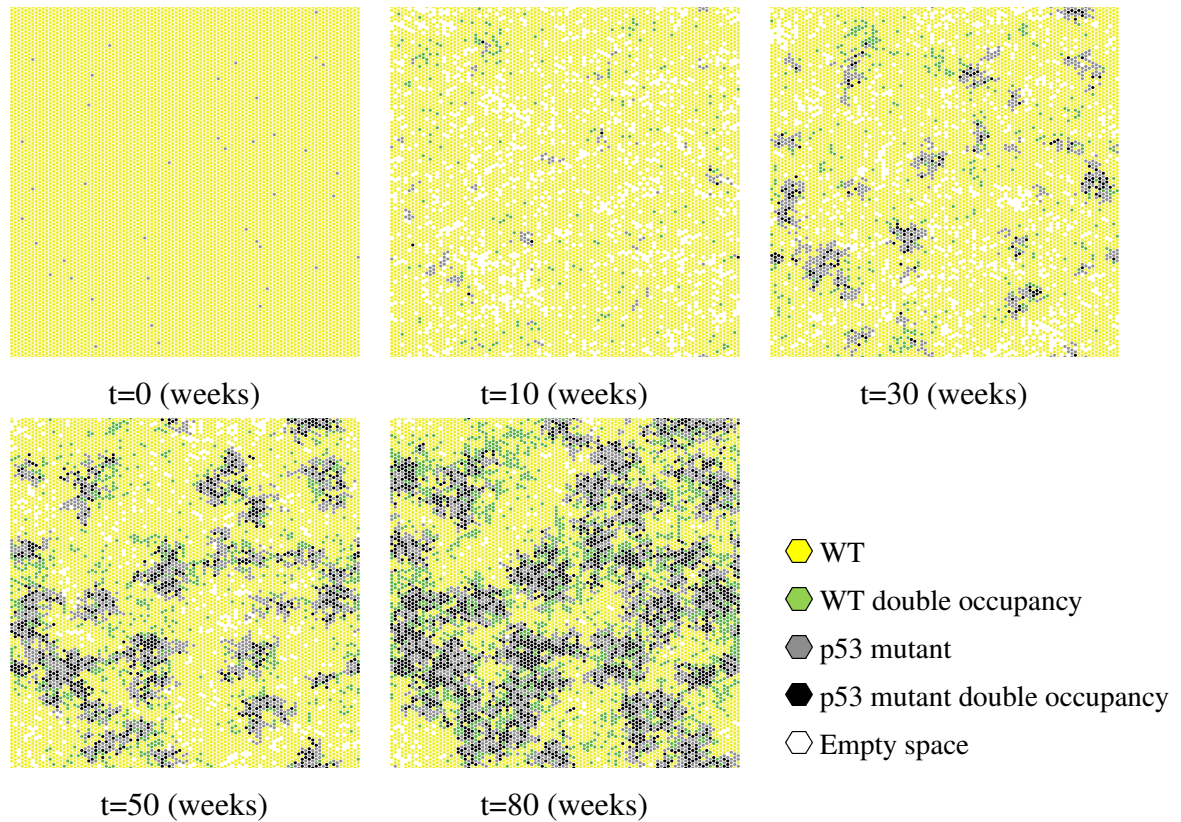
Fig. 5.21 Combined downstream feedback and crowding release time lapse. Typical simulation time lapse of the SP spatial model of p53 mutant growth with combined downstream feedbacks and crowding release. Cells are placed on a 2D 100x100 hexagonal lattice. The combined feedback model is in quantitative agreement with the experimental observations and successfully maintains tissue turnover.

(a)

(b)

(c)

(d)
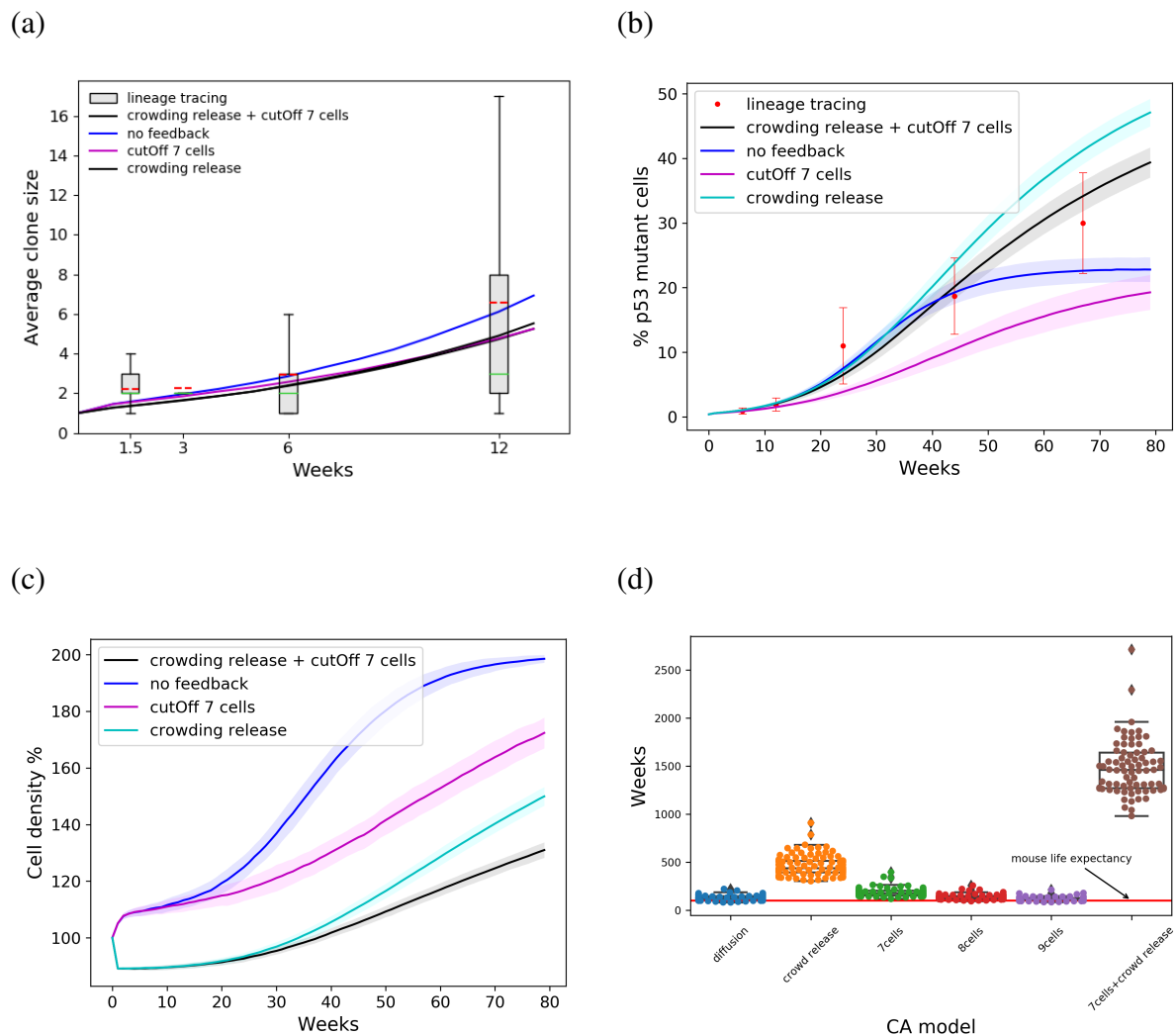
Fig. 5.22 Comparison of different spatial feedback models simulating p53 mutant growth in the basal layer of back epidermis. Parameters: $r = 0.06$, $\rho = 0.77$, $\lambda = 1.16$/week, $\delta = 0.95$, from Murai et al. (2018). The combined model describes the data more appropriately. a) Simulated p53 average clone size, box plots correspond to experimental clone counts, dashed red lines: mean, solid green line: median, b) simulated and observed percentage of p53 mutant cells, c) simulated overall basal cell density, d) last simulated time point of each repetition (100 simulation repetitions were performed). a,b,c: Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

The combined feedback model is in quantitative agreement with the experimentally observed average clone size (Figure 5.22a) and tissue takeover (Figure 5.22b). Moreover a less crowded basal cell density is produced (Figure 5.22c). Importantly, such model is also able to simulate tissue turnover for substantially larger periods, rendering potential tissue dynamics simulations of a human system feasible (Figure 5.22d).

Table 5.1 provide provides a summary of all simulation setups performed to test possible feedback regimes and identify which one appropriately describes p53 mutant dynamics.

Setups	Mechanical		Cell-Cell Communication		Key Observations
	Diffusion	Crowding Release	Up-stream	Down-stream	
1	✓				Mutants spread faster, overcrowding.
2		✓			No overcrowding, Takeover overestimation at later time points.
3			✓		Enhanced takeover, no fit to experiments.
4				✓	No overcrowding, takeover underestimation, Overcrowding at high crowding cut-offs
5		✓		✓	**No overcrowding**, **agreement with experiments**

Table 5.1 Feedback simulation setups. A summary of the computational experiments performed for identifying the most appropriate feedback mechanism that explains the behaviour of $p53^{*/wt}$ cells in skin. Both mechanical and cell-cell communication feedback mechanisms explored. A downstream feedback mechanism combined with crowding release successfully reproduces p53 mutant dynamics.

Considering the above, it is evident that a downstream feedback mechanism with instant stratification of double cells and a relaxed crowding cut-off can suitably explain p53 mutant behaviour in mouse epidermis.

5.5 Spatial simulations of p53 clone growth in oesophageal epithelium

Having specified the spatial rules for the downstream feedback mechanisms which succeed in both recapitulating p53 mutant dynamics and maintaining tissue survival in back epidermis, I then sought to validate whether such feedback model would be suitable to describe p53 mutant behaviour in an independent dataset. To this end, I used a mouse oesophageal p53 mutant clonal dataset, provided by Phil Jones at Sanger Institute.

To estimate the parameters r, λ, ρ and δ for the mutant simulations, I assumed that r, λ and ρ remain unchanged, given that $p53^{*/wt}$ cells divide and stratify at the same rates as their wild type counterparts. Based on this assumption the only unknown parameter would be the fate bias of mutant cells, δ. From Klein et al. (2010), δ can be estimated from a linear regression applied to the ln(average clone size), as shown in Figure 5.23. The slope of the line equals $2\delta r\lambda$, where δ is the unknown parameter.
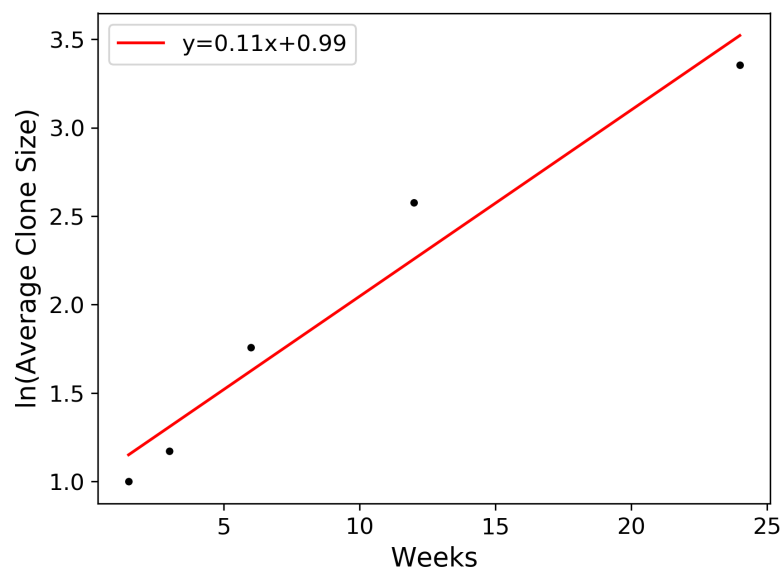
Fig. 5.23 Estimate δ for the oesophageal dataset. Linear regression applied to the ln(average clone size) of $p53^{*/wt}$ cells in oesophageal epithelium over all time points. The slope of the line gives $2\delta r\lambda$.

The r, λ and ρ parameters describing the homeostatic cell population dynamics in oesophageal epithelium are provided by Doupé et al. (2012). However, as also specified in 4.5, page 102 two more parameter combinations are available, inferred from quantitative clonal analysis of alternative oesophageal datasets. The additional parameter sets resulted from the following steps: new H2B-GFP dilution experiments using a more sensitive microscope were performed. Cell cycle properties estimated from these experiments were incorporated in model simulations and new parameter values were measured using a maximum likelihood inference approach (Piedrafita et al., 2020). With that in mind, simulations of the downstream feedback model in oesophageal epithelium were performed using all three available parameter sets proposed for this tissue (Table 5.2).

Analysis of the simulated p53 mutant cell populations and clone sizes demonstrates that the model is in quantitative agreement with the experimental values (Figure 5.24). This confirms that the downstream feedback mechanism is appropriate to describe the behaviour

Tissue	Reference	λ(/week)	r	ρ	δ
Oesophagus (OE)	Doupé et al. (2012)	1.9	0.1	0.65	0.29
Oesophagus (OE)	Piedrafita et al. (2020)	2.9	0.06	0.56	0.32
Oesophagus (OE)	Piedrafita et al. (2020)	2.9	0.1	0.65	0.19

Table 5.2 SP model parameter sets for oesophageal epithelium. Historical (Doupé et al., 2012) and recent (Piedrafita et al., 2020) parameter values inferred for progenitor cell behaviour in murine oesophageal epithelium, as derived from quantitative lineage tracing. Fate bias parameter (δ) was estimated from linear regression applied to the ln(average clone size) of $p53^{*/wt}$ cells.

of p53 mutant cells in the oesophageal epithelium and is also largely robust to different parameter sets inferred for this tissue.

Taken together, the growth of p53 mutant cells resembles a logistic curve and is shown to be described in space by a non neutral SP model with cell density feedback rules applied downstream to p53 mutations. Importantly, such spatial model recapitulates the observed mutant behaviour in datasets from two different tissues, the mouse oesophagus and back epidermis. The downstream feedback mechanism is consistent with the p53 mutants' adaptable fate, suggested by their observed incomplete tissue takeover.

(a)

(b)

(c)
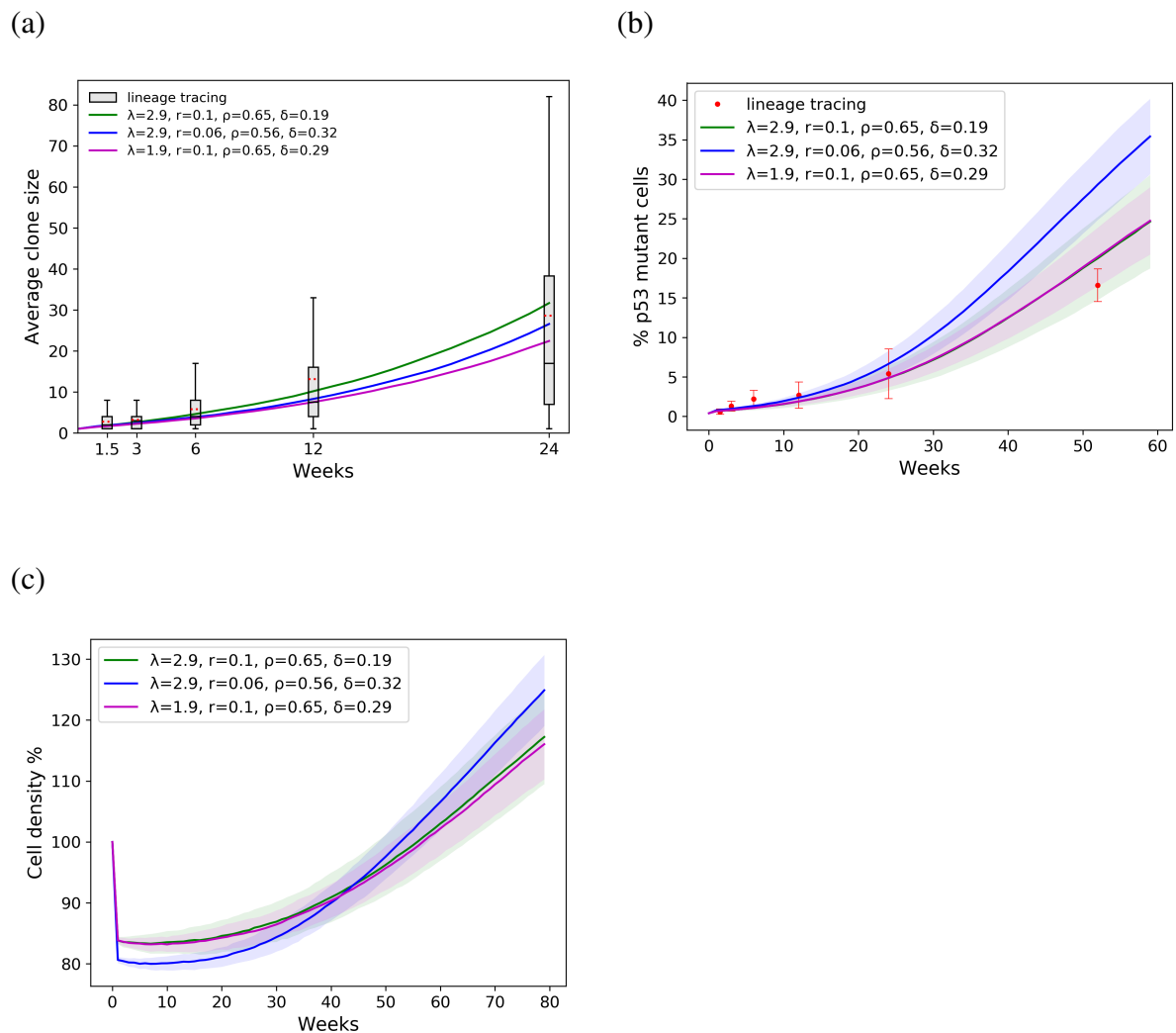
Fig. 5.24 Quantitative analysis of the spatial fate bias SP model with the downstream feedback, simulating p53 mutant growth in the basal layer of oesophageal epithelium under different inferred parameter sets. a) Simulated p53 average clone size, box plots correspond to experimental clone counts, dashed red lines: mean, solid green line: median, b) simulated and observed percentage of p53 mutant cells, c) simulated overall basal cell density. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

5.6   Loss of Maml maintains clone advantage through a uniform response to crowding

I have already demonstrated through spatial modelling and analysis of p53 mutant growth in mouse back skin and oesophagus that a downstream feedback model explains the behaviour of this type of mutation. I then sought to model how Maml mutants evolve and compete in space as they are shown to exhibit a more aggressive behaviour compared to $p53^{*/wt}$ cells.

Specifically, early dynamics of Maml knock outs have been shown to have their symmetric differentiation probabilities minimized and promote the differentiation of adjacent wild type cells (Alcolea et al., 2014). Moreover, they proliferate 3 fold faster and stratify at lower rates. As a consequence, at 52 weeks Maml clones colonized the entire epithelium causing a $\sim 10\%$ increase in the basal cell density. However, tissue integrity was not compromised as no tumours were formed (Alcolea et al., 2014).

To answer the question whether tissue colonization by Maml clones can be attributed to their fate imbalance and different division and stratification rates alone, I used the initial spatial non neutral SP model without feedbacks ((5.2), page 115). Similarly to p53 mutations, Maml simulations were set up and performed as described in Section 4.3, page 80. The grid was initialized by randomly seeded proliferating and differentiating basal cells, with 1% of the lattice sites being occupied by mutants to mimic the reported induction of Maml cells in the corresponding lineage tracing experiment. The following parameter values from Alcolea et al. (2014), were used:

Maml cells: $r = 0.055$, $\rho = 0.12$, $\lambda = 6.0\ week^{-1}$, $\Gamma = 0.8\ week^{-1}$, $\delta = 1.0$.
WT cells: $r = 0.1$, $\rho = 0.65$, $\lambda = 1.9\ week^{-1}$, $\Gamma = 3.5\ week^{-1}$, $\delta = 0$.

5.6.1   Spatial simulations of Maml clone growth in OE

Spatial simulations of Maml cells growing in a wild type background using the spatial non neutral SP model were not able to reproduce Maml mutant behaviour (Figure 5.25). Maml cells are able to occupy only 30% of the basal layer (Figure 5.25,a). A 100% increase in the basal layer density is also observed (Figure 5.25,b), highlighting the already mentioned issue of overcrowding and the need for introducing spatial feedback rules.

I then investigated whether the downstream feedback model which was previously shown to successfully describe p53 mutant behaviour would be able to reproduce the dynamics of Maml mutants. Spatial simulations of the downstream feedback model did not achieve the experimentally observed full tissue takeover at one year post induction (Figure 5.26,a). There is an enhanced expansion in Maml cell populations compared to the p53 equivalent simulations, which could be attributed to their faster division and slower stratification rates. Nevertheless, even the apparent advantage of an elevated proliferation rate does not make the Maml cells able to occupy the entire grid. Moreover, the inclusion of the downstream feedback rules is not able to prevent the overcrowding effect (Figure 5.26,b).

(a)   (b)
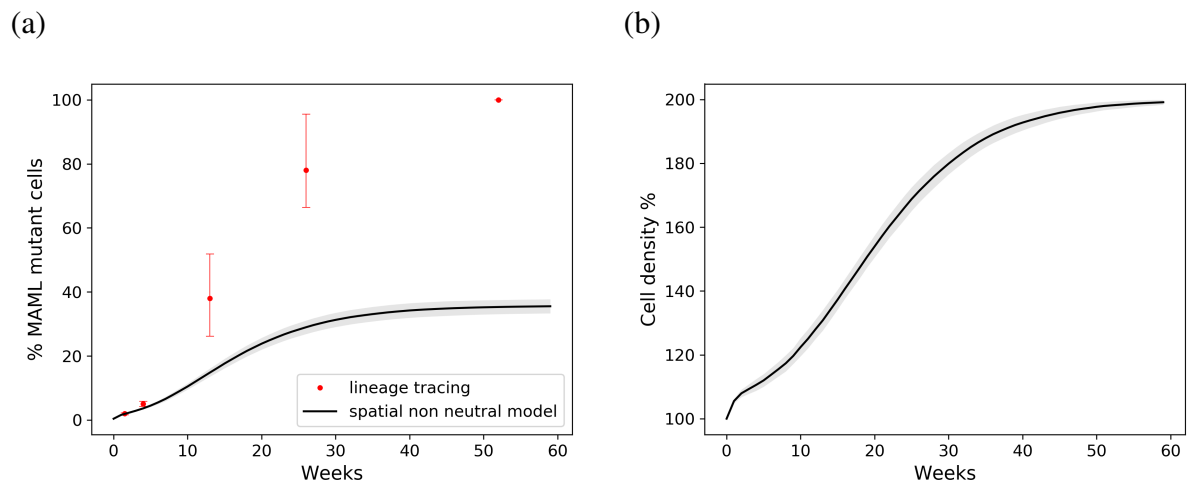
Fig. 5.25 Quantitative analysis of the spatial fate bias SP model, simulating Maml mutant growth in the basal layer of oesophageal epithelium. Parameters used from Alcolea et al. (2014) a) simulated and observed percentage of Maml mutant cells disagree, b) simulated overall basal cell density increases substantially, reflecting overcrowding issues. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

(a)   (b)
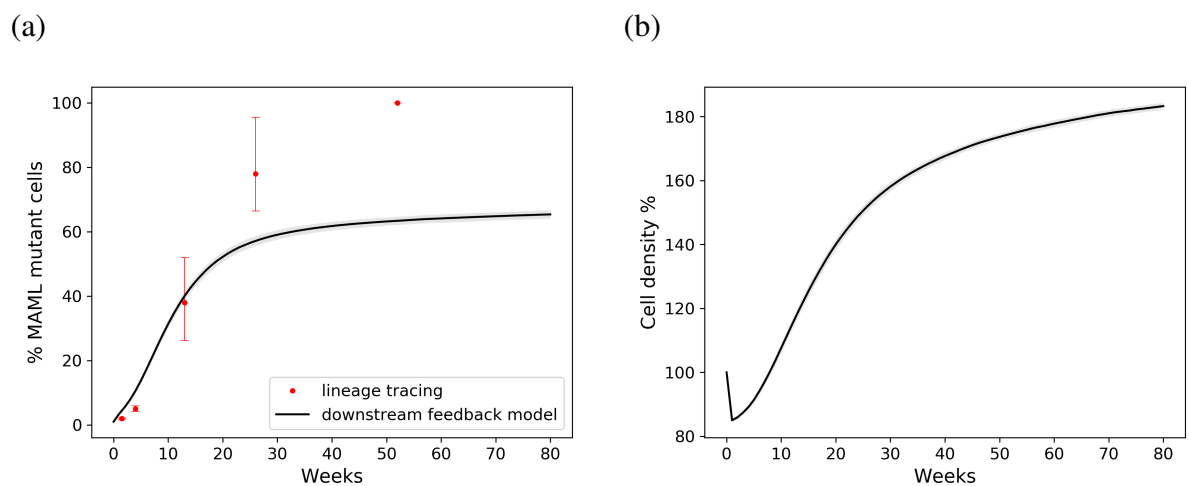
Fig. 5.26 Quantitative analysis of the spatial fate bias SP model with downstream feedback, simulating Maml mutant growth in the basal layer of oesophageal epithelium. Parameters used from Alcolea et al. (2014) a) simulated and observed percentage of Maml mutant cells disagree as simulated Maml population only partially colonise the tissue, b) simulated overall basal cell density increases. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

The disagreement between computational simulations and experimental observations implies that the growth of Maml mutants is distinct to that of $p53^{*/wt}$ cells. This leads to the rejection of the downstream feedback model and the seeking of an alternative feedback mechanism to explain Maml mutant behaviour. The ability of these mutations to fully take

over the tissue would be more compatible with a feedback that affects both mutant and wild-type cells equally. Considering this, I performed simulations of the upstream response rule (5.4.2, page 133), where the cell density feedback is applied to all cell populations.
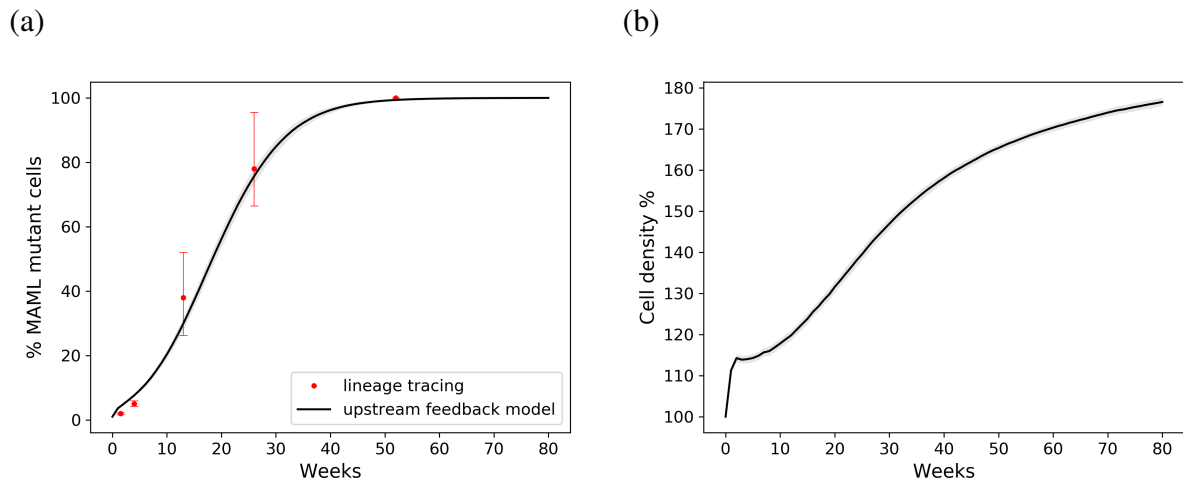
Fig. 5.27 Quantitative analysis of the spatial fate bias SP model with upstream feedback, simulating Maml mutant growth in the basal layer of oesophageal epithelium. Parameters used from Alcolea et al. (2014). Density feedback bias ($\delta' = 1$) and crowding cut-off was set to 7 cells. a) Tissue takeover by Maml cell population is successfully reproduced, b) simulated overall basal cell density increases. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.
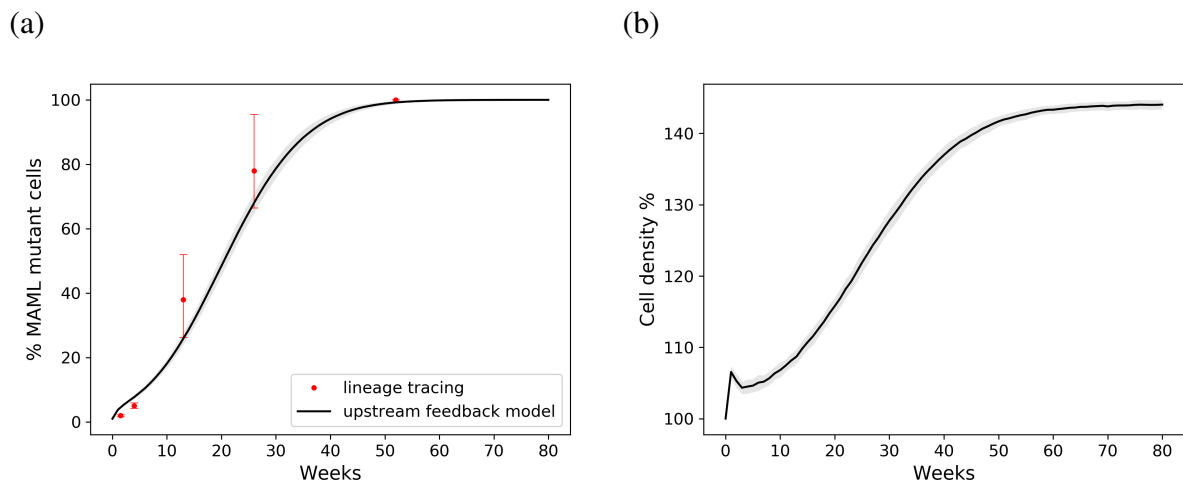
Fig. 5.28 Quantitative analysis of the spatial fate bias SP model with upstream feedback, setting the probability of symmetric division to zero in the case of an overcrowded neighbourhood. Parameters used from Alcolea et al. (2014). Density feedback bias ($\delta' = 1$) and crowding cut-off was set to 6 cells. a) modelled percentage of Maml mutants fit to observations, b) cell density increases to a lower extent. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

Simulating Maml spatial growth using upstream feedbacks gave an excellent fit to the experimental observations of tissue takeover (Figure 5.27a), indicating that this type of spatial rules can suitably explain Maml behaviour. However, there was a 60% increase in cell density at the end of 52 weeks, which contradicts the experimentally observed $\sim 10\%$ (Figure 5.27,b). To address this issue, I repeated the simulations by adding an extra rule which was applied to every cell in the grid that had an overcrowded neighbourhood ($n > 8$ cells). In such case the probability of that cell to undergo symmetric division was minimized.

Visual examination of the underlying simulations suggested that the difference in growth patterns between the upstream and downstream models arises as in the upstream model the reduction in δ in all cells had the effect on ensuring that cells continued dividing, but maintained the relative advantage of the mutant clone. This effect, where Maml1 loss gives a constant fitness advantage whilst the p53 mutant advantage is transient, was subsequently confirmed using a spatial Moran model.
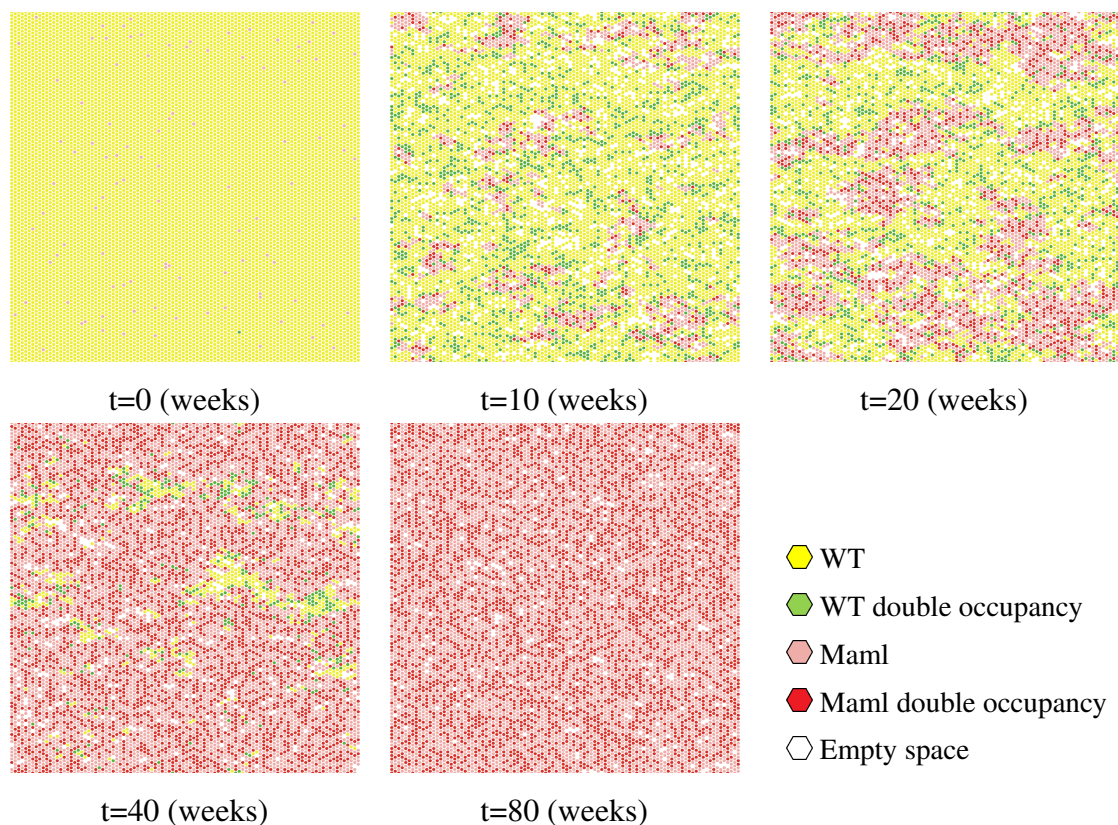
Fig. 5.29 Upstream model time lapse simulating Maml population. Typical simulation time lapse of the SP spatial model of Maml mutant growth with upstream feedback. Cells are placed on a 2D 100x100 hexagonal lattice. Maml cells takeover the whole grid, successfully recapitulating experimental observations.

The new simulations presented a slightly compromised fit in the percentage of Maml cells (Figure 5.28a) but led to a more moderate increase in basal cell density ($\sim 40\%$) by 52 weeks (Figure 5.28b). As mutant behaviour was still largely reproduced, that model could be considered as appropriate for describing Maml dynamics. This is also confirmed by visual examination of the underlying simulations (Figure 5.29). The upstream feedback mechanism successfully reproduced the growth pattern of Maml cells, as the response to crowding by both wild type and mutant populations had the effect on ensuring that cells continued dividing, but maintained the relative advantage of the mutant clones.

5.7 p53 and Maml in competition

In the previous sections of this chapter, I explored how two different mutant types (p53 and Maml) grow and compete for space within a wild type environment. To do this, I used spatial computational simulations which were compared with existing experimental lineage tracing datasets in mouse stratified squamous epithelial tissues. I demonstrated that the logistic growth of $p53^{*/wt}$ cells can be described by a model with density-dependent feedbacks activated downstream to p53 mutations, therefore affecting solely mutant cells. Conversely, I found that a model with feedbacks affecting both wild type and mutant cell populations is more appropriate to describe the evolution of Maml knock out clones. These results imply that the two mutant populations show a diverse response to crowding, providing a mechanistic explanation of the observed distinct growth modes.

Having shown that both sets of mutations can be understood in terms of the interactions of the underlying pathways and the tissue, I sought to assess how a single model that could enable both feedbacks behaved and how the mutations coexist and interact in the tissue. For that purpose, I implemented a unified spatial model, where I introduced downstream and upstream feedback rules in order to be able to model both mutants in the same simulation. Clonal data from healthy tissues and tumours has shown that in squamous oesophageal tumours, p53 is relatively enriched in the tumour, but *NOTCH1* mutations are depleted (Martincorena et al., 2015, 2018), strongly suggesting that in healthy tissue *NOTCH1* loss has a tumour protective effect (Higa and DeGregori, 2019). As Maml represents a gene with control over several Notch related genes, I therefore modelled the effect of *NOTCH1* and p53 mutations in order to examine how the different growth behaviours in the tissue alter the competition. Mutant competition analysis was performed for oesophageal epithelium.

(a)   (b)
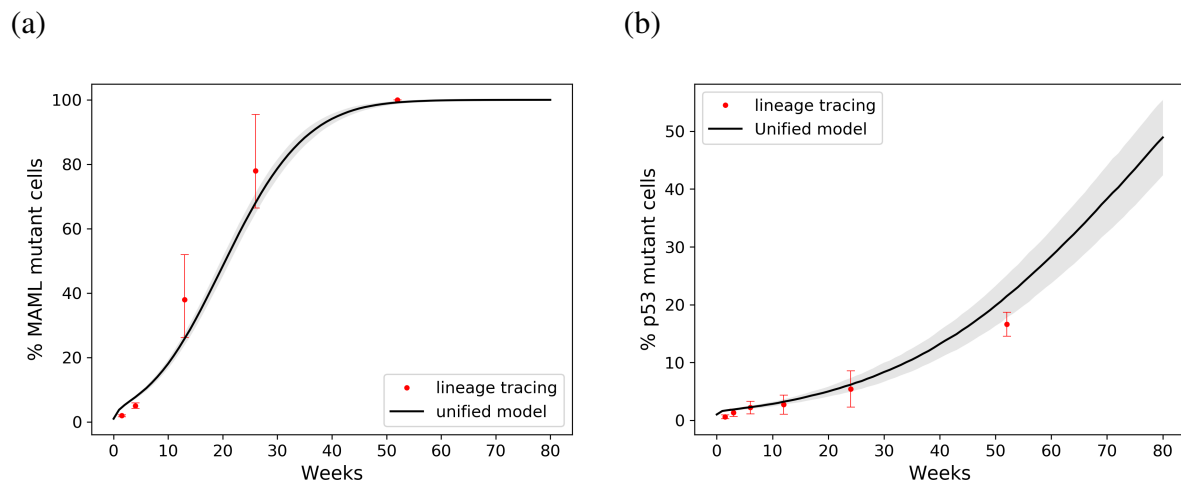
Fig. 5.30 Simulations of the unified model with p53 and Maml mutants separately match experimental data. a) simulated and observed percentage of Maml mutant cells in OE, b) simulated and observed percentage of p53 mutant cells in OE. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

To ensure the validity of the unified model, I performed simulations of p53 and Maml mutants separately and confirmed that the individual behaviour of each mutant type was successfully reproduced (Figure 5.30).

The following parameter values were used:

p53 cells: $r = 0.1$, $\rho = 0.65$, $\lambda = 1.9\ week^{-1}$, $\Gamma = 3.5\ week^{-1}$, $\delta = 0.29$.

Maml cells: $r = 0.055$, $\rho = 0.12$, $\lambda = 6.0\ week^{-1}$, $\Gamma = 0.8\ week^{-1}$, $\delta = 1.0$.

WT cells: $r = 0.1$, $\rho = 0.65$, $\lambda = 1.9\ week^{-1}$, $\Gamma = 3.5\ week^{-1}$, $\delta = 0$.

I then proceeded to perform simulations of Maml and p53 mutants competing in a wild type background. The grid was initialized by randomly seeded proliferating and differentiating basal cells, with 1% of the lattice sites being occupied by p53 mutants and another 1% by Maml mutants. A striking effect resulting from mutant clone spatial interaction is that $p53^{*/wt}$ cell population appears to shrink (Figure 5.31).
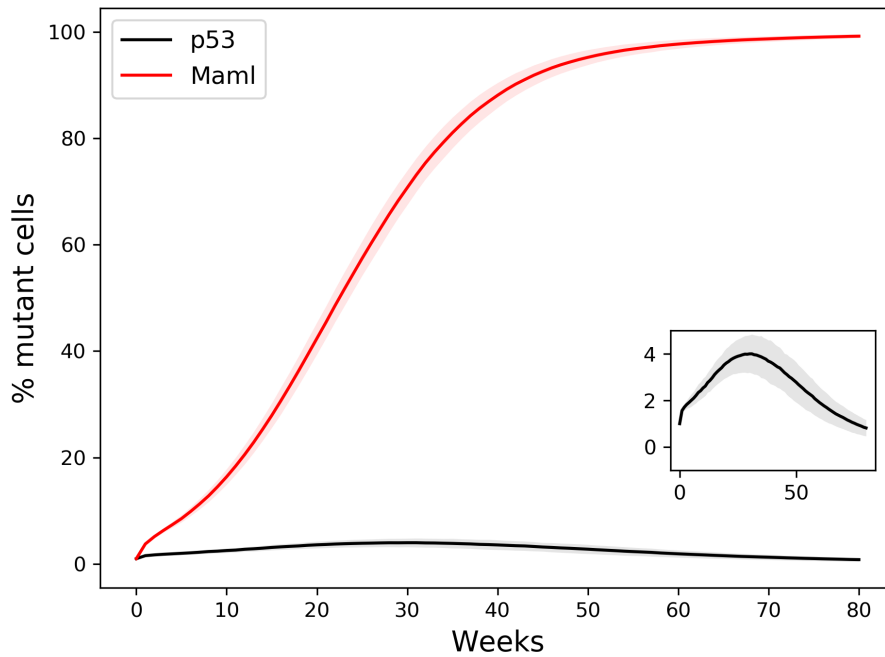
Fig. 5.31 Competition simulations of p53 and Maml mutant cells. p53 and Maml mutants were co-induced in a grid with wild type cells to investigate their spatial interaction. Maml mutants eventually took over the tissue, outcompeting p53 mutant population. p53 mutant shrinkage is easier observed in the inset plot, which shows the takeover of p53 mutant population on a smaller scale.

To gain insights into the effect of mutant competition in space, I sought to quantitatively describe potential different spatial features of the two mutant populations. To this end, I used the following metrics for each mutant population:

- proportion of boundary cells: cells that have at least one neighbour of a different type.

$$\frac{b_m}{c_m},$$

  where b_m denotes the number of boundary cells for mutation m and c_m the total number of cells for mutation m.

- cell mixing index: the proportion of neighbours belonging to a different type

$$\frac{\sum_{i=1}^{b_m} \frac{n_i'}{n_i}}{c_m},$$

  where n_i' denotes the number of different type neighbouring cells of a boundary cell i, and n_i the total number of neighbours of boundary cell i.

The above metrics were measured for all p53 and Maml cells at different time points (Figure 5.32). From the outputs of this analysis, it is evident that the two different mutations show distinct properties. The majority of p53 mutant cells are consistently found in boundaries with Maml or wild type populations, implying that $p53^{*/wt}$ cells are not able to form coherent groups in space (Figure 5.32a). On the contrary, the proportion of boundary Maml cells progressively decreases and is almost minimized at later time points, as they colonize the entire grid. Furthermore, p53 mutants mixing index is consistently higher, indicating that they are more likely to share junctions with different cell types and therefore are more dispersed (Figure 5.32b).

(a)   (b)
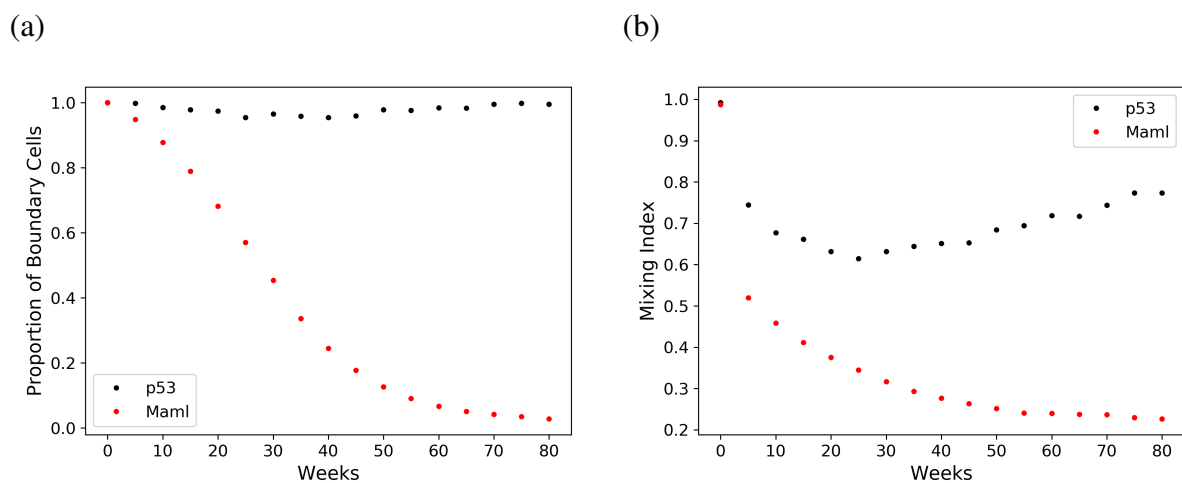
Fig. 5.32 Quantitative analysis of mutant spatial features. a) Proportion of boundary cells in p53 and Maml cells, b) cell mixing index, which is proportion of neighbours belonging to a different type. The cell mixing index was calculated for each mutant cell and the values were averaged over each mutation.

In seeking to understand more aspects of p53 and Maml mutant competition in space, I designed a series of different mutational events and explored how these affected the behaviour of the two mutant populations. The different competition scenarios included co-induction, timed inductions, and variable levels of relative induction.

(a)

(b)

(c)

(d)

(e)

(f)
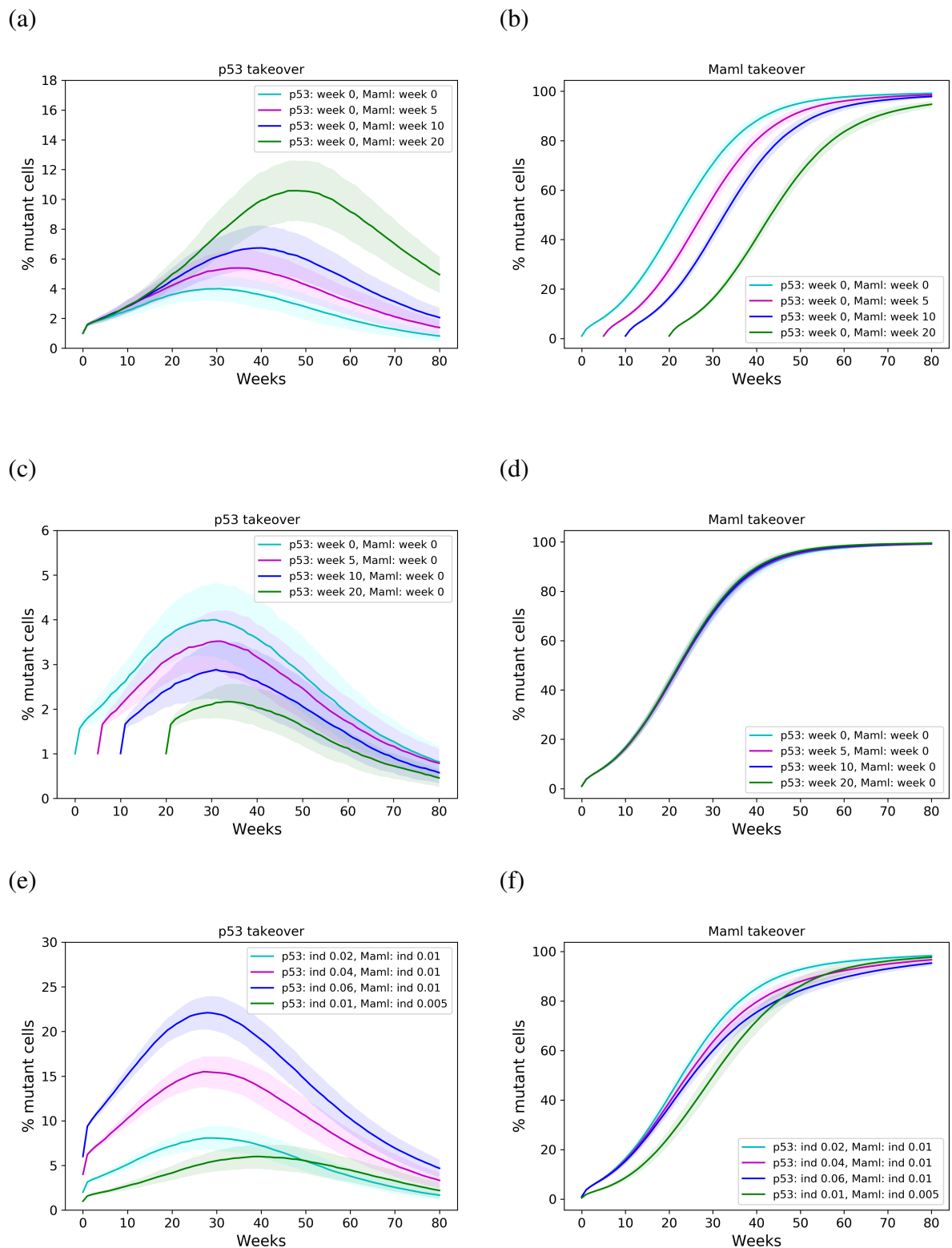
Fig. 5.33 Tissue takeover of p53 and Maml mutant populations across different competition scenarios where either the induction time or the induction level were changed. p53 (a) and Maml (b) tissue takeover in simulations where Maml cells were introduced at 0, 5, 10 and 20 weeks, p53 (c) and Maml (d) tissue takeover in simulations where p53 cells were introduced at 0, 5, 10 and 20 weeks, p53 (e) and Maml (f) tissue takeover in simulations where the two mutant populations were introduced in different proportions in the tissue. Data correspond to mean values across 100 simulations. Shaded areas correspond to SD.

Strikingly, in every simulation scenario Maml population outcompetes $p53^{*/wt}$. p53 mutant clones that had grown for extended periods before Maml mutations were introduced regressed more slowly, as the larger clones were more slowly outcompeted but ultimately the constant fitness advantage offered by Maml mutations eventually led to the loss of $p53^{*/wt}$ clones. (Figure 5.33). As expected, the later the Maml population is introduced, the higher the percentage of p53 mutant population reaches in the tissue (Figure 5.33a,b). In $p53^{*/wt}$ delayed introduction simulations, the later the p53 mutants are introduced the more difficult is for them to expand as Maml cells have time to grow without competition and are already established when p53 mutant cells are inserted to the system (Figure 5.33c,d). Similar dynamics is observed in settings of variable induction levels. Simulations where p53 mutant induction was increased twofold, fourfold and sixfold were performed as well as scenarios where Maml induction was reduced by half were tested. As expected, an increased takeover of p53 mutants is observed when their induction level is higher but soon their percentage drops (Figure 5.33e,f). Collectively, the above observations highlight that Maml knock outs appear consistently more aggressive over their p53 competitors.

Aside from the apparent "winner" status of Maml population, the outcomes of the different competing settings underline interesting aspects of the "loser", p53 mutant behaviour. A closer inspection of the p53 mutant takeover plots reveals that the rate at which $p53^{*/wt}$ takeover diminishes varies across different cases (Figure 5.33). For instance, one can see that the highest proportions of p53 mutant population in the tissue are achieved in the simulations where p53 induction level is increased sixfold or fourfold (Figure 5.33e). Furthermore, an increased p53 takeover is observed in the case where Maml cells are introduced at the 20th week (Figure 5.33a). Despite the enhanced tissue takeover percentage at higher induction levels, its drop appears to be quicker compared to the delayed Maml induction condition.

To further investigate that, I assessed the loss rate of p53 mutant population across all competition scenarios. Figure 5.34 presents the average percentage loss in $p53^{*/wt}$ tissue takeover per week, starting from the week when takeover was the highest until the final week of the simulation. Interestingly, these calculations confirm that the p53 takeover loss rate is higher when the competition simulation is initialized with a substantially higher proportion of $p53^{*/wt}$ compared to Maml cells.
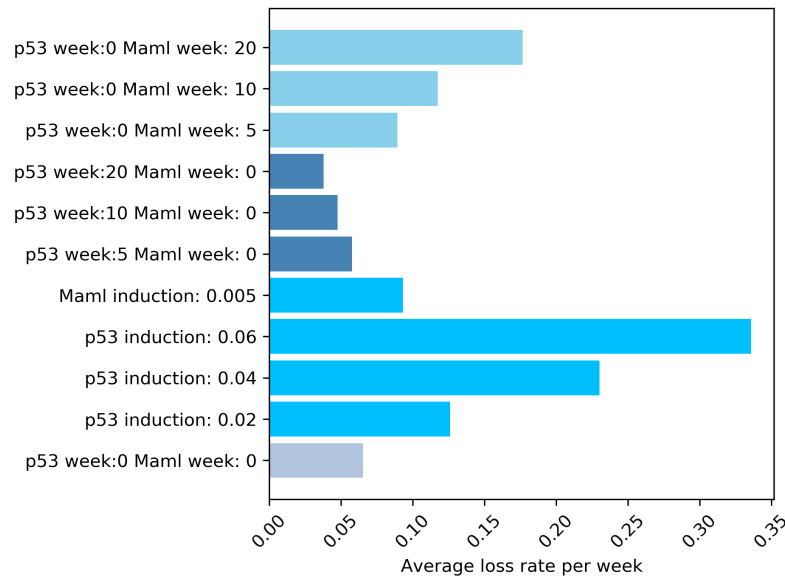
Fig. 5.34 $p53^{*/wt}$ **loss rate.** Average percentage of p53 mutant takeover loss rate across all simulated competition scenarios and the base competition simulations (same induction level, same induction time point). Calculations measured the difference in tissue takeover between the week when takeover was the highest until the final week of the simulation.

The above findings are also compatible with the analysis of the boundary cell proportion and mixing index. In simulations of Maml induction at 20 weeks, the proportion of p53 boundary cells and the p53 mixing index are reaching their minimum values between 40 to 50 weeks (Figure 5.35a,b), concurring with the time period when $p53^{*/wt}$ takeover is reaching its peak (Figure 5.33a). An analogous pattern is observed in the spatial metrics of simulations with increased p53 induction level. The peak week of tissue takeover (Figure 5.33e) coincides with the time point where the proportion of boundary cells and the mixing index of p53 mutants decrease (Figure 5.35c,d). The lower values of the spatial quantitative metrics indicate that p53 mutant cells appear less dispersed and form more coherent clusters when their tissue takeover reaches its maximum level.

A direct comparison of the spatial metrics values between these two competition settings highlights a shifting in the axis of time (Figure 5.35e,f). When p53 mutant induction is increased, both boundary cell proportion and mixing index reach their minimum values several weeks earlier compared to the scenario where Maml mutants are introduced at 20 weeks ($\sim$ 30 weeks vs $\sim$ 50 weeks respectively). Following their minimum value, the two spatial metrics start rising again, indicating an increase in the level of cell dispersion. This increase appears to happen earlier when p53 mutants are introduced in increased numbers, consistent with the higher loss rate observed for this simulation condition (Figure 5.34).
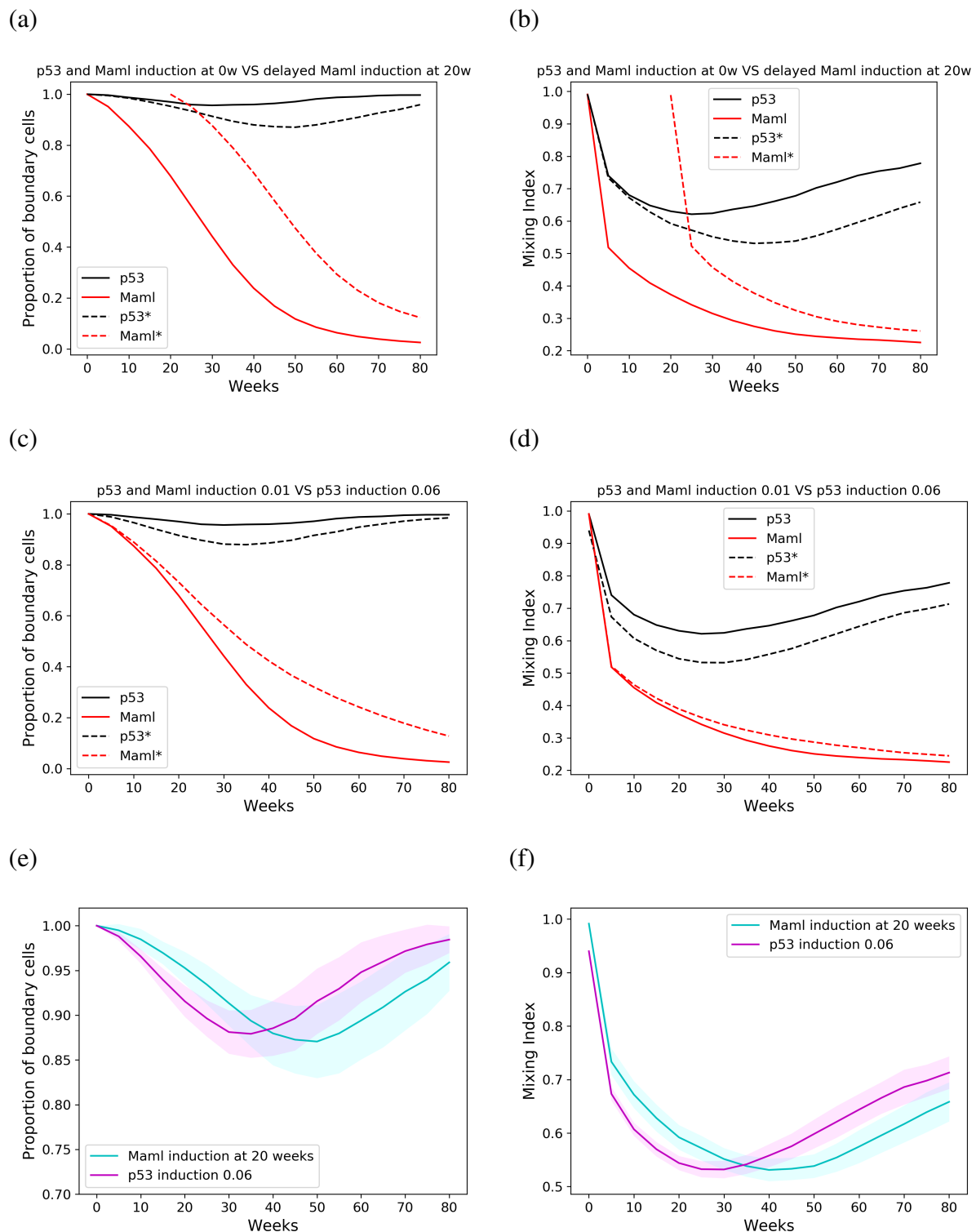
Fig. 5.35 Proportion of boundary cells and proportion of different type neighbours (cell mixing index) in Maml and p53 mutants as calculated for different competition scenarios. a,b) simulations of **p53** cells introduced at **0 week** at 0.01 induction and **Maml** cells introduced at **20 weeks** at 0.01 induction, c,d) simulations of **p53** cells introduced at 0 week at **0.06 induction** and **Maml** cells introduced at 0 week at **0.01 induction**, e,f) comparison of the two scenarios. a,b,c,d: solid black line corresponds calculations for the p53 population at the base competition simulations (same induction level, same induction time point), black dashed line corresponds to calculations for the p53 population at the given scenario, red solid line corresponds to calculations for the Maml population at the base competition simulations (same induction level, same induction time point), red dashed line corresponds to calculations for the Maml population at the given scenario. Data to correspond to mean values across 100 simulations.

A visual inspection of the simulations considering two above mentioned competition scenarios can give more insights in the way the two mutant populations interact in space. When p53 mutants are introduced in the grid in much higher numbers compared to their Maml competitors (Figures 5.36, 5.37), they would expand and form aggregates more easily and faster. However, as they cluster in space and tend to appear in more crowded environments and encounter more $p53^{*/wt}$ neighbours, they would start losing their innate proliferating advantage. The subsequent contact with Maml cells which start expanding substantially in the following time points would lead to their loss as the Maml population is more aggressive.

t=10 (weeks)   t=30 (weeks)   t=60 (weeks)

◯ WT   ◯ Maml   ⬡ Maml double occupancy   ⬡ p53   ⬢ p53 double occupancy
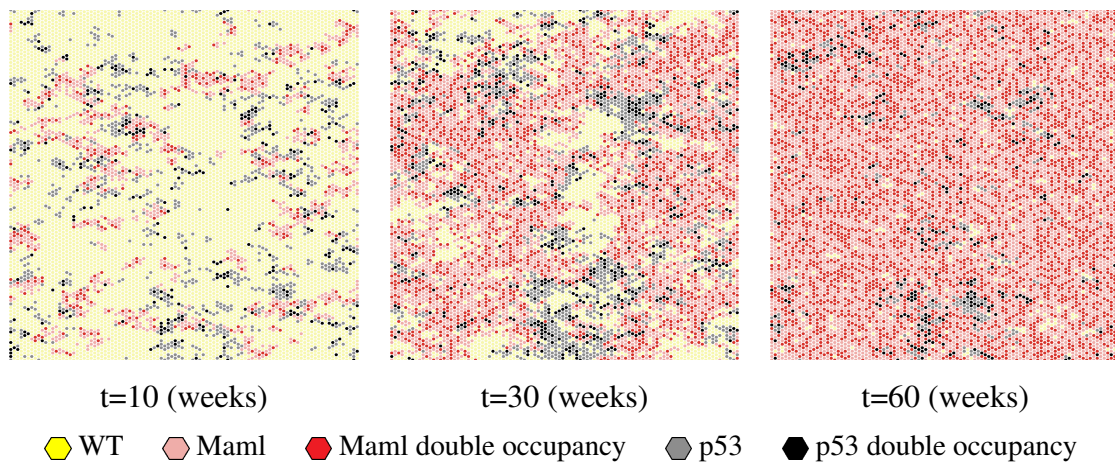
Fig. 5.36 Enhanced p53 mutant induction scenario. Typical simulation time lapse of the SP spatial competition model with enhanced p53 mutant induction level. Cells are placed on a 2D 100x100 hexagonal lattice. $p53^{*/wt}$ cells are introduced in proportions six fold higher than Maml cells (0.06 and 0.01 respectively).

t=20 (weeks)   t=45 (weeks)   t=80 (weeks)

◯ WT   ◯ Maml   ⬡ Maml double occupancy   ⬡ p53   ⬢ p53 double occupancy
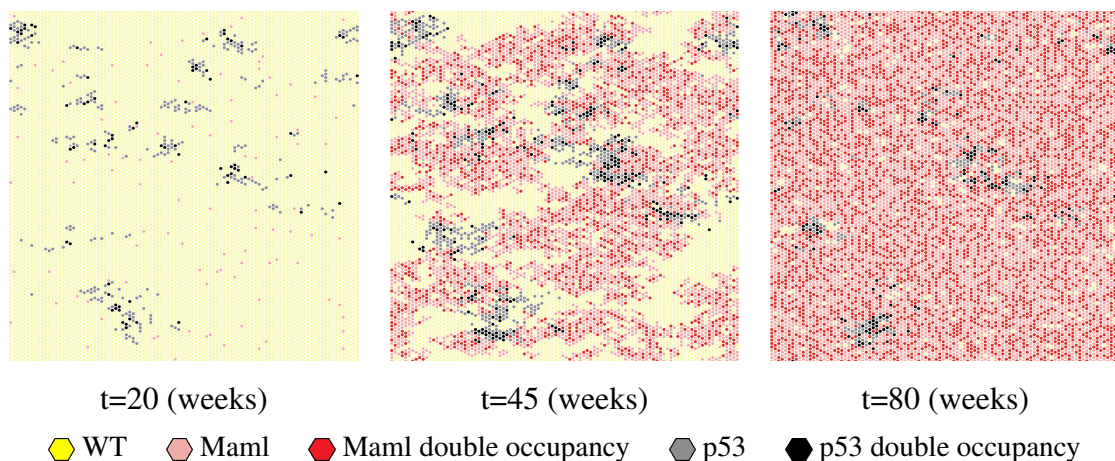
Fig. 5.37 Delayed Maml induction scenario. Typical simulation time lapse of the SP spatial competition model with delayed Maml induction time. Cells are placed on a 2D 100x100 hexagonal lattice. $p53^{*/wt}$ cells are introduced at the beginning of the simulation whereas Maml cells at 20 weeks.

5.8   Cell competition as a wider paradigm

Competition simulations of p53 and Maml mutants might suggest that pre-cancer dynamics can be explained in terms of clone competition within the tissue environment. That it to say, the way cells interact in space dictates growth dynamics and may determines how mutations expand and whether this limits or promotes cancer development. The competition hypothesis can be used to provide a conceptual explanation on the observed dynamics in patients with oesophageal cancer. Barrett's oesophagus is a pre-malignant state, where normal squamous epithelial stem cells in the lower oesophagus are transformed to simple columnar type cells. It has been observed that patients with tumours that are adjacent to Barrett's tissue have a higher survival, though the mechanism for this is as yet unclear (Sawas et al., 2018). The increased survival may be a result of competition between populations of cells, as the presence of Barrett's tissue effectively restrains the growth of the tumour. Therefore, the ultimate fate of Barrett's tissue may be determined by mutations received early in the development process. That is to say, whether a region of Barrett's tissue will develop into cancer may be determined by mutations that occur prior to Barrett's expansion and presentation in the clinic. To test this hypothesis explicitly I performed simulations of the spatial SP model considering two configurations:

1. Two cell clones with competitive advantage over wild type cells, growing next to each other, which represented the case of tumour adjacent Barrett's. Barrett's cells were considered to be less fit than tumour cells.

2. One cell clone growing in a wild type environment, which represented the case of tumour with no adjacent Barrett's.

Both mixed and single populations were simulated in a space originally containing 10,000 cells, and each simulation was repeated 100 times. One can observe that in competition with Barrett's cells, cancer cells are effectively restrained in their growth and take over a smaller portion of the tissue (Figure 5.38). In the absence of a competing Barrett's population, tumour cells take over most of the space. Together, these observations are consistent with the hypothesis that competition, even with a less fit population, can restrain the growth of the tumour and thereby increase survival. Although these models do have simplifications, they demonstrate a general principle of how spatial competition influences cancer progression.

(a)
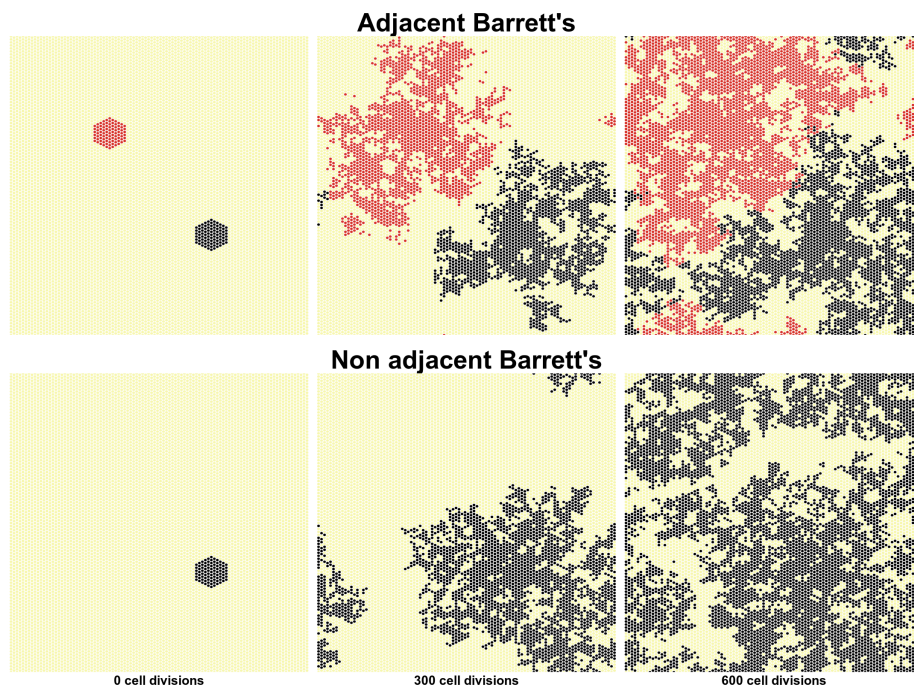
(b)
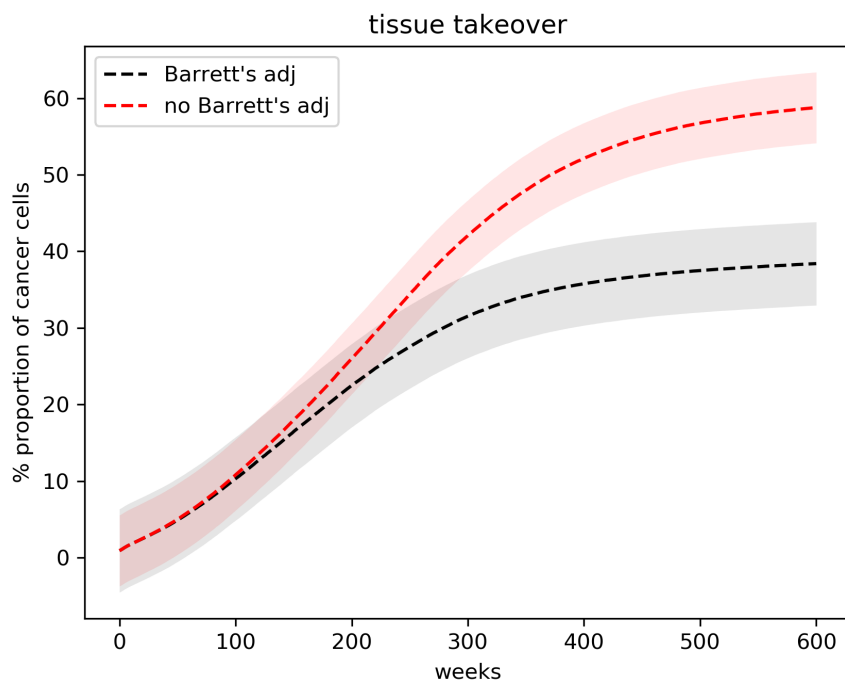
Fig. 5.38 Adjacent Barrett's oesophagus limits the growth of cancer clones, revealed by cellular automata simulations. a) Visualisations of the growth of cancer clones (black) in a predominantly wild type environment (yellow) shows how takeover is effectively limited by the presence of Barrett's clones (red). b) Averaging takeover across multiple repetitions highlights that this process is highly robust.

5.9 Wounding

The introduction of feedbacks to the system further allows testing for response to tissue level challenges. Wounding events drive proliferation to close wounds behind a non-proliferating front (Rognoni and Watt, 2018, Doupé et al., 2012). Whilst the spatial SP model does not describe the peripheral cells, I performed simulations with wounding and compared the wound closure behaviour with experimental data. Several recent studies have investigated the process of wound repair through computational modelling (Staddon et al., 2018, Tetley et al., 2019, Andasari et al., 2018, Li et al., 2016).

To investigate wound healing with the spatial SP model, I induced wounding events by introducing a patch of vacant lattice sites in the grid and simulated wound closure. Two simulation approaches were followed: in the first approach, I included feedbacks in the model (cells responded to local density as in the upstream feedback mechanism) whilst in the second, no feedbacks were applied. I found that when feedbacks are included, the gap in the grid is closed at a faster rate, showing that the model with feedbacks is more efficient at wound healing (Figure 5.39). However, the spatial model is substantially slower at closing the wound compared to experimental data from mouse oesophagus, which report that the tissue was repaired in 5 days (Doupé et al., 2012). This suggests that the cell migration observed in tissue wounding situations needs to be explicitly included in the spatial model to simulate tissue repair, in order to achieve a faster closure rate comparable to the one observed experimentally.
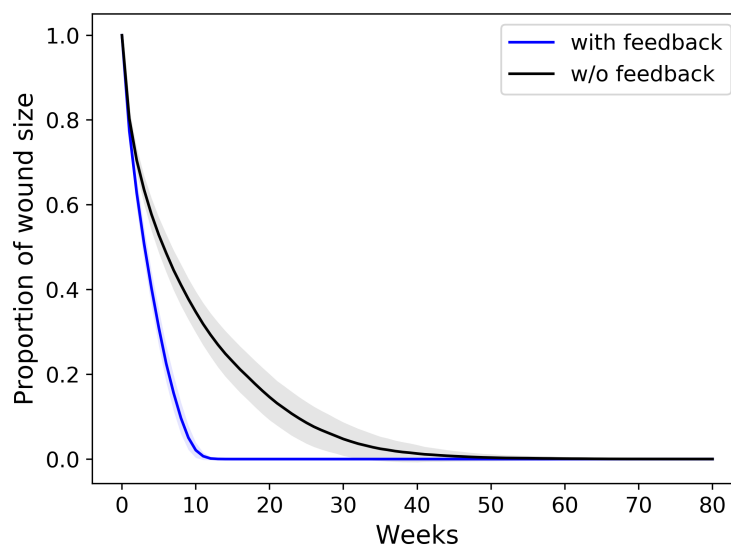
Fig. 5.39 Simulating wound healing. The SP spatial model with feedbacks closes wounds at a faster rate compared to the non-feedback model.

5.10   Discussion

To summarize, in this chapter I explored the growth and competition of non-neutral mutations in stratified squamous epithelia using spatial models. Mutations in the tumour suppressor gene p53 and inhibition of the Notch pathway through Maml knock out were modelled. The studied dynamics of p53 and Maml clones in mouse epithelia has been found to be highly distinct. Whilst Maml knock outs are able to dominate the whole tissue (Alcolea et al., 2014), p53 mutants initially expand but they eventually occupy no more than 30% of the tissue (Murai et al., 2018). I showed that in order to recapitulate these behaviours whilst maintaining tissue turnover the spatial models need to take account of feedbacks between neighbouring cells in the tissue.

I demonstrated that p53 clone growth approximates a logistic curve, but that without including limitations on mutation induced expansion the overall proliferation rate of the tissue drops due to space restrictions. In contrast, the ability of Maml mutations to take over the stem cell population reflects a feedback that affects both mutant and wild-type cells equally. Thus, for reproducing p53 mutant behaviour, I used a model where the response to crowding events was applied solely to mutant cells, indicating activation downstream to p53 mutations ("downstream" feedback). Such model appears consistent with the p53 mutants' adaptable fate, suggested by their observed incomplete tissue takeover. On the contrary, I modelled Maml dynamics using a feedback mechanism which required an independent or upstream to mutation response to crowding ("upstream" feedback), consistent with the ability of Maml cells to evict their neighbours during their early growth (Alcolea et al., 2014). These spatial models of Maml and p53 mutant growth could provide a mechanistic explanation of the observed distinct behaviours of the two mutant types, suggesting an increased sensitivity of $p53^{*/wt}$ clones to cell density.

p53 and Maml competition in space was also explored. A striking effect resulting from the spatial interaction of the two mutations in a wild type background is that $p53^{*/wt}$ cell population is outcompeted by Maml population and appears to shrink. The winning behaviour of Maml cells is consistent across a range of competition scenarios, where the two mutant populations are introduced in the simulations at different time points or proportions. A further quantitative comparison of the spatial features of mutant competition revealed distinct growth patterns. Whilst Maml cells form coherent aggregates in space which progressively expand at the expense of their neighbours, $p53^{*/wt}$ cells initially cluster but eventually appear more dispersed and less cohesive.

Taken together, these findings suggest that the distinct tissue takeover outcomes may be attributed to the differential response of p53 and Maml mutations to crowding. Moreover, according to the competition simulations, the sensitivity of p53 mutant cells to increased cell density environments might be the cause of their loser status. The crowding sensitivity as a hallmark of loser cells is also supported by previous studies (Eisenhoffer et al., 2012, Marinari et al., 2012, Wagstaff et al., 2016, Levayer et al., 2016). Furthermore, the increased dispersion of p53 mutants at later time points in competition simulations, indicate that this spatial emergent property might be a result of their mixing with Maml populations. This would be consistent with previous studies, correlating the probability of loser cell elimination with the surface of contact shared with winners (Levayer et al., 2015, Bove et al., 2017). Further experimental work on p53 and Maml competition in mouse epithelia would shed more light on how these two mutants interact in mammalian tissues.

The two distinct feedbacks mechanisms that describe mutant dynamics may suggest the distribution of mutations observed in human datasets. Recent studies identified high incidence of many of the frequently found mutated genes in tumours such as *TP53* and *NOTCH1* in healthy human skin and oesophagus (Martincorena et al., 2015, 2018, Yokoyama et al., 2019). Strikingly, a higher frequency of *NOTCH1* mutations is reported for healthy human oesophagus compared to oesophageal squamous cell carcinomas whilst the opposite applies to *TP53* mutations. Considering the consistent winning behaviour of *NOTCH1* over p53 mutants in competition simulations and given the paucity of *NOTCH1* mutations in tumours, it is tempting to speculate that the aggressive fitness of *NOTCH1* may offer a tumour-protective effect.

Chapter 6

Discussion

Abstract

In this thesis, I explored how space changes growth and how it affects the ability of mutant cells to spread within tissues, providing useful insights on understanding pre-cancer dynamics. Towards this goal, I initially sought to quantify epithelial homeostatic dynamics and presented a method that appropriately parameterises the system.

I have demonstrated that the explicit inclusion of spatial constraints do not alter neutral dynamics in squamous epithelia. However, non-neutral dynamics require the incorporation of feedbacks based on the response to the neighbouring cellular density. Based on such feedbacks, I investigated the growth of advantageous mutations and provided a mechanistic explanation of the observed differences in their growth.

A particular outcome of this thesis is that spatial competition and the way mutant cells respond to spatial constraints defines the early steps of tumour formation.

6.1   Parameterisation of epithelial homeostasis

Epithelial tissue maintenance is particularly important to understand homeostasis and processes such as aging, preneoplasia and cancer formation. The investigation of mutant and pre-neoplastic dynamics necessitates understanding how a healthy tissue is maintained. Therefore, it is important to retrieve quantitative information of the homeostatic epithelial stem cell dynamics through accurate parameterisation, which was my initial focus. Stem cell dynamics in multiple squamous epithelial tissue has been argued over several years to be described by the single progenitor model (Clayton et al., 2007, Doupé et al., 2010,

2012, Lim et al., 2013, Rompolas et al., 2016). Using three parameters, the division rate, the stratification rate, and the probability of symmetric division, this model predicts average clone size, clone size distributions, tissue homeostasis, and cell survival probabilities.

In this thesis, I sought to test and validate the SP model in depth. I identified issues in the proposed analytical solution for SP parameter inference, with the most important one being the collision in parameter probabilities. I further tested the analytical approach on synthetic datasets and found that the method was overly precise, producing unrealistic 95% confidence intervals. Importantly, when tested synthetic datasets that explicitly accounted for realistic cell cycle distributions and inter-mice biological variation, the proposed parameter values were inaccurate with the true values outside error bars. I explored and compared alternative strategies for analyzing experimental datasets, identifying an SMC-ABC approach as the best in terms of efficiency and appropriate error estimation.

The main outcome of the above analysis was the identification of the most appropriate method for analysing lineage tracing data which would allow new datasets to be analysed accurately and efficiently. The above findings showed that the parameter inference process is highly sensitive to biological variation and highlighted the need to be extremely cautious with statistical analysis especially when sampling is low. This is further demonstrated by the analysis of synthetic datasets with included inter-mice variability. These datasets were generated assuming a typical number of mice used in a lineage tracing experiment (two or three per time point). The likelihood distributions calculated from individual time points varied across time and were insufficient to estimate parameters on their own. When compared with likelihood distributions calculated from experimental measurements, whilst the distributions were generally highly similar, the small number of biological replicates at each time point leaves them prone to distortion by chance parameter combinations. Given that the appearance of individual time points has been used to justify the consideration of more complex models of stem cell hierarchy (Mascré et al., 2012), this could influence the interpretation of experimental data analysis. These findings would suggest that increasing sampling at individual time points whilst reducing the total number of time points would minimize this effect without impacting parameter estimation. Considering that the experimental protocols using animal models are expensive and time consuming, this would be a recommendation for future experimental design.

6.2 Neutral dynamics is unchanged when space is explicitly considered

Having identified a suitable method for accurate parameterisation, I further challenged the SP model and introduced spatial constraints to explore how these altered neutral dynamics. Simulations of the spatial SP model revealed that the model continues to describe neutral growth as it was able to replicate the key homeostatic signatures on stem cell population dynamics. This was consistent across different datasets. Nevertheless, the spatial model generated some interesting emergent properties. As cell fate decision is considered a stochastic, cell intrinsic process, local imbalances in cell density (gaps or thickenings) were observed. These events were transient but not realistic as demonstrated by comparison with microscopy tissue images. The presence of unrealistic gaps or double cells might reflect the spatial restrictions imposed by the underlying cellular automaton grid, as cells were restricted in a fixed regular lattice. Visualising the grid differently, e.g. through Voronoi diagrams might overcome such issues. Despite the emergence of these properties, the spatial SP model successfully recovered the homeostatic behaviour.

The variability in cell density observed in simulations suggested that *in vivo*, cell fate decision outcomes might be directly coupled to keep a homeostatic balance. The coupling of division and stratification is a concept already discussed in previous studies. Recent work of Miroshnikova and colleagues suggested that proliferation drives local stratification in mouse developing epidermis (Miroshnikova et al., 2018). Furthermore Mesa et al. (2018) proposed that stratification triggers the division of nearby cells in adult mouse plantar epidermis. However, variations of the spatial SP model where division and stratification events were linked did not seem to offer a substantial improvement to the expected homeostatic dynamics. On the contrary, the stratification driven regime produced a high proportion of empty sites, thus failing to reproduce homeostasis for most of the parameter sets tested. This indicates that this model might be appropriate for certain parameter sets. Additionally, the division driven model revealed patterns of clustering in proliferating and differentiating cells.

Collectively, the above observations demonstrate how small changes in a simple model may lead to interesting patterns of packing. Considering the different proposed models for epithelial homeostasis, it would be interesting to test whether a more complex, hierarchical model would be successful in recapitulating the signatures of homeostasis if implemented in space. The dynamics of a simple model may be also altered by the explicit consideration of the characteristic tissue geometry. That is to say, the underlying structural properties of different tissues have a direct impact on cell growth and spatial distribution. For instance, in

tissues with a more complex architecture such as the intestinal crypts, the compartmentalised spatial organisation of cells into glands leads to segregated cell populations and limits the extent to which individual clones can grow. Quantifying the spatial patterns in cell organisation and clone shapes using spatial metrics would allow further assessment of the effects of alternative models and different geometries on dynamics. This could be addressed in future research.

6.3 Competition of advantageous mutations in space

In order to explore competition of advantageous mutations within the tissue, I challenged the spatial SP model even further by considering non-neutral growth. The role of space is of paramount importance, dictating how mutants accumulate, interact and spread in tissues. Competition between cells in tissues determines the ability of individual clones to dominate the tissue. I modelled the dynamics of p53 and Notch related (Maml1) mutations, which have been shown experimentally to grow in distinct ways (Murai et al., 2018, Alcolea et al., 2014). In contrast to the spatial simulations of homeostasis, I found that the spatial SP model was unable to simulate non-neutral dynamics without accounting for local cell density dependent feedbacks. Therefore, I modelled non-neutral growth considering a spatial SP model with both a growth advantage but also innate negative feedbacks based on cellular density. Interestingly, the introduction of such feedbacks in simulations of the neutral system fixed the previously observed cell density fluctuations, as the local cell density distribution was comparable to that in microscopy images.

Based on observed growth patterns, I proposed two distinct feedback mechanisms that successfully reproduce the observed p53 and Notch behaviour, thus providing a mechanistic explanation of the different growth patterns. Consistent with their rapid takeover and ability to eject their neighbours during early growth as shown by EdU labelling experiments (Alcolea et al., 2014), I modelled Maml1 growth considering an upstream to mutations response to crowding ("upstream" feedback). On the contrary, the transient advantage of p53 mutant clones suggested a feedback response downstream to p53 mutations ("downstream" feedback). The simulations of p53 mutant growth showed that when p53 clones collide they lose their advantage and their growth becomes slowed, suggesting a differential sensitivity of the two mutant types to crowding. These feedbacks were based on the cellular density (i.e. number of cells) in each cell's neighbourhood. Alternative density metrics, e.g. the proportion of proliferating cells in each cell's neighbourhood could also be tested.

As p53 and Maml1 mutants presented a highly distinct behaviour, it would be interesting to investigate whether this is reflected in differences in the morphologies of their cells and/or clones. Therefore, an interesting future direction would involve the comparative analysis of cell and clone shapes between different mutations and the wild type population. The lineage tracing data available did not provide information regarding physical properties of the system, such as stress magnitude. New experiments measuring cell mechanics would assist in simulating and analysing these properties using more complex cell based models which incorporate cell mechanical parameters explicitly.

The aggressive phenotype of Maml1 cells was maintained in competition simulations. Maml1 mutants always acted as "winners", outcompeting p53 mutant cells, a behaviour consistent across a series of different competition settings. Together, these findings may have implications on the distribution of mutations observed in human datasets. Although mutations in *TP53* and *NOTCH1* genes are considered cancer causing, it has been shown recently that *TP53* is relatively enriched in squamous oesophageal tumours whereas *NOTCH1* mutations are depleted. Additionally, the same studies report a higher and lower frequency of *NOTCH1* and *TP53* mutations respectively in phenotypically healthy tissues (Martincorena et al., 2015, 2018, Yokoyama et al., 2019). These observations, along with the consistent winning behaviour of *NOTCH1* over p53 mutants in competition simulations strongly suggest that non-neutral competition may offer a tumour protective effect.

This hypothesis demonstrates that pre-cancer dynamics and progression to cancer can be explained in terms of clone competition within the tissue environment. The way cells interact in space dictates growth dynamics and may determines how mutations expand and whether this limits or promotes cancer development. Spatial competition could also be considered as a paradigm for understanding the progression of mutated cells from Barrett's to cancer. The increased survival observed in patients with tumours growing adjacent to Barrett's tissue (Sawas et al., 2018) may be a result of competition between populations of cells, as the presence of Barrett's tissue effectively restrains the growth of the tumour.

The work presented in this thesis highlighted the implications of cell competition within a spatially constrained tissue environment in pre-cancer progression. As spatial modelling has been shown to provide useful insights on tissue dynamics, cell competition in space could be further explored to answer new questions. The lattice models used for the purposes of this thesis allowed me to identify the mechanisms of cell growth and competition within the spatial constraints of epithelial tissues. The work presented here can be further extended by exploring new aspects of cell competition outcomes using lattice free models. In particular, a lattice free approach would enable the study and analysis of cell/clone shape and cell size

alterations as it allows for a more realistic representation of cell physical properties. This in turn, would permit a more detailed analysis of areas of cell contacts, cell movement as well as types of cell spatial topologies and interactions. These properties may be important parameters in providing more clear evidence on how cell populations interact, how mutants spread and how losers are eliminated.

One of the main outcomes of this work is the effects of cell crowding on tissue dynamics and the need of incorporating feedbacks based on the response to local cell density. As an off-lattice approach would allow for a more realistic representation of local cell crowding, it would be therefore an interesting future research direction to study the effects of mechanical cell competition by quantifying cell density differences and mechanical strain in a lattice free setting.

The outputs of all the above analyses could provide new testable hypotheses which can then be combined with new experimental data to infer the role of tissue geometry and spatial properties of the local environment on growth dynamics.

References

Abe, Y. and Tanaka, N. (2017). Roles of the hedgehog signaling pathway in epidermal and hair follicle development, homeostasis, and cancer. *Journal of Developmental Biology*, 5(4):12.

Agrawal, N., Jiao, Y., Bettegowda, C., Hutfless, S. M., Wang, Y., David, S., Cheng, Y., Twaddell, W. S., Latt, N. L., Shin, E. J., Wang, L.-D., Wang, L., Yang, W., Velculescu, V. E., Vogelstein, B., Papadopoulos, N., Kinzler, K. W., and Meltzer, S. J. (2012). Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. *Cancer Discovery*, 2(10):899–905.

Alcolea, M. P. (2017). Oesophageal stem cells and cancer. In Birbrair, A., editor, *Advances in Experimental Medicine and Biology*, volume 1041, pages 187–206. Springer International Publishing.

Alcolea, M. P., Greulich, P., Wabik, A., Frede, J., Simons, B. D., and Jones, P. H. (2014). Differentiation imbalance in single oesophageal progenitor cells causes clonal immortalization and field change. *Nature Cell Biology*, 16(6):612–619.

Alcolea, M. P. and Jones, P. H. (2013). Tracking cells in their native habitat: lineage tracing in epithelial neoplasia. *Nature Reviews Cancer*, 13(3):161–171.

Alcolea, M. P. and Jones, P. H. (2014). Lineage analysis of epidermal stem cells. *Cold Spring Harbor Perspectives in Medicine*, 4(1):1–15.

Andasari, V., Lü, D., Swat, M., Feng, S., Spill, F., Chen, L., Luo, X., Zaman, M., and Long, M. (2018). Computational model of wound healing: EGF secreted by fibroblasts promotes delayed re-epithelialization of epithelial keratinocytes. *Integrative Biology*, 10(10):605–634.

Antal, T. and Krapivsky, P. L. (2010). Exact solution of a two-type branching process: clone size distribution in cell division kinetics. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(07):P07028.

Banks, H., Broido, A., Canter, B., Gayvert, K., Hu, S., Joyner, M., and Link, K. (2012). Simulation algorithms for continuous time markov chain models. *Studies in Applied Electromagnetics and Mechanics*, 37:3–18.

Barrandon, Y. and Green, H. (1987). Three clonal types of keratinocyte with different capacities for multiplication. *Proceedings of the National Academy of Sciences*, 84(8):2302–2306.

Beaumont, M. A. (2010). Approximate bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41(1):379–406.

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate bayesian computation in population genetics. *Genetics*, 162(4):2025–2035.

Bielmeier, C., Alt, S., Weichselberger, V., La Fortezza, M., Harz, H., Jülicher, F., Salbreux, G., and Classen, A.-K. (2016). Interface contractility between differently fated cells drives cell elimination and cyst formation. *Current Biology*, 26(5):563 – 574.

Blanpain, C. and Fuchs, E. (2009). Epidermal homeostasis: a balancing act of stem cells in the skin. *Nature Reviews Molecular Cell Biology*, 10(3):207–217.

Blanpain, C. and Simons, B. D. (2013). Unravelling stem cell dynamics by lineage tracing. *Nature Reviews Molecular Cell Biology*, 14(8):489–502.

Block, M., Schöll, E., and Drasdo, D. (2007). Classifying the expansion kinetics and critical surface dynamics of growing cell populations. *Physical Review Letters*, 99(24):248101.

Blum, M. G. (2010). Choosing the summary statistics and the acceptance rate in approximate bayesian computation. In Lechevallier, Y. and Saporta, G., editors, *Proceedings of COMPSTAT'2010*, pages 47–56. Physica-Verlag HD.

Bove, A., Gradeci, D., Fujita, Y., Banerjee, S., Charras, G., and Lowe, A. R. (2017). Local cellular neighborhood controls proliferation in cell competition. *Molecular Biology of the Cell*, 28(23):3215–3228.

Candi, E., Schmidt, R., and Melino, G. (2005). The cornified envelope: a model of cell death in the skin. *Nature Reviews Molecular Cell Biology*, 6(4):328–340.

Chamseddine, I. M. and Rejniak, K. A. (2019). Hybrid modeling frameworks of tumor development and treatment. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 12(1):e1461.

Chkhaidze, K., Heide, T., Werner, B., Williams, M. J., Huang, W., Caravagna, G., Graham, T. A., and Sottoriva, A. (2019). Spatially constrained tumour growth affects the patterns of clonal selection and neutral drift in cancer genomic data. *PLOS Computational Biology*, 15(7):e1007243.

Ciocchetta, F. and Hillston, J. (2009). Bio-PEPA: A framework for the modelling and analysis of biological systems. *Theoretical Computer Science*, 410(33-34):3065–3084.

Clayton, E., Doupé, D. P., Klein, A. M., Winton, D. J., Simons, B. D., and Jones, P. H. (2007). A single type of progenitor cell maintains normal epidermis. *Nature*, 446(7132):185–9.

Croagh, D., Phillips, W. A., Redvers, R., Thomas, R. J., and Kaur, P. (2007). Identification of candidate murine esophageal stem cells using a combination of cell kinetic studies and cell surface markers. *Stem Cells*, 25(2):313–318.

Deutsch, A. (2005). *Cellular automaton modeling of biological pattern formation : characterization, applications, and analysis*. Birkhauser, Boston.

Donati, G. and Watt, F. M. (2015). Stem cell heterogeneity and plasticity in epithelia. *Cell Stem Cell*, 16(5):465–476.

Dotto, G. P. and Rustgi, A. K. (2016). Squamous cell cancers: A unified perspective on biology and genetics. *Cancer Cell*, 29(5):622–637.

Doupé, D. P. (2009). *Quantitative analysis of epithelial homeostasis*. PhD thesis, University of Cambridge.

Doupé, D. P., Alcolea, M. P., Roshan, A., Zhang, G., Klein, A. M., Simons, B. D., and Jones, P. H. (2012). A single progenitor population switches behavior to maintain and repair esophageal epithelium. *Science*, 337(6098):1091–1093.

Doupé, D. P. and Jones, P. H. (2013). Cycling progenitors maintain epithelia while diverse cell types contribute to repair. *BioEssays*, 35(5):443–451.

Doupé, D. P., Klein, A. M., Simons, B. D., and Jones, P. H. (2010). The ordered architecture of murine ear epidermis is maintained by progenitor cells with random fate. *Developmental Cell*, 18(2):317–323.

Efron, B. and Stein, C. (1981). The jackknife estimate of variance. *The Annals of Statistics*, 9(3):586–596.

Eisenhoffer, G. T., Loftus, P. D., Yoshigi, M., Otsuna, H., Chien, C. B., Morcos, P. A., and Rosenblatt, J. (2012). Crowding induces live cell extrusion to maintain homeostatic cell numbers in epithelia. *Nature*, 484(7395):546–549.

Erban, R., Chapman, J., and Maini, P. (2007). A practical guide to stochastic simulations of reaction-diffusion processes. *arXiv preprint arXiv:0704.1908*, page 35.

Faisal, M., Futschik, A., and Hussain, I. (2013). A new approach to choose acceptance cutoff for approximate bayesian computation. *Journal of Applied Statistics*, 40(4):862–869.

Filippi, S., Barnes, C. P., Cornebise, J., and Stumpf, M. P. (2013). On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo. *Statistical Applications in Genetics and Molecular Biology*, 12(1):87–107.

Fletcher, A. G., Osterfield, M., Baker, R. E., and Shvartsman, S. Y. (2014). Vertex models of epithelial morphogenesis. *Biophysical Journal*, 106(11):2291–2304.

Frank, S. (2007). Stem cells: tissue renewal. In *Dynamics of Cancer: Incidence, Inheritance, and Evolution*, chapter 12. Princeton University Press.

Frede, J., Greulich, P., Nagy, T., Simons, B. D., and Jones, P. H. (2016). A single dividing cell population with imbalanced fate drives oesophageal tumour growth. *Nature Cell Biology*, 18(9):967–978.

Gandini, S., Sera, F., Cattaruzza, M. S., Pasquini, P., Picconi, O., Boyle, P., and Melchi, C. F. (2005). Meta-analysis of risk factors for cutaneous melanoma: II. sun exposure. *European Journal of Cancer*, 41(1):45–60.

Gao, Y.-B., Chen, Z.-L., Li, J.-G., Hu, X.-D., Shi, X.-J., Sun, Z.-M., Zhang, F., Zhao, Z.-R., Li, Z.-T., Liu, Z.-Y., Zhao, Y.-D., Sun, J., Zhou, C.-C., Yao, R., Wang, S.-Y., Wang, P., Sun, N., Zhang, B.-H., Dong, J.-S., Yu, Y., Luo, M., Feng, X.-L., Shi, S.-S., Zhou, F., Tan, F.-W., Qiu, B., Li, N., Shao, K., Zhang, L.-J., Zhang, L.-J., Xue, Q., Gao, S.-G., and He, J. (2014). Genetic landscape of esophageal squamous cell carcinoma. *Nature Genetics*, 46(10):1097–1102.

Gardner, M. (1970). Mathematical games. *Scientific American*, 223(4):120–123.

Gevertz, J. L. and Torquato, S. (2006). Modeling the effects of vasculature evolution on early brain tumor growth. *Journal of Theoretical Biology*, 243(4):517 – 531.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361.

Gillespie, D. T. (2001). Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733.

Graner, F. and Glazier, J. A. (1992). Simulation of biological cell sorting using a two-dimensional extended potts model. *Physical Review Letters*, 69(13):2013–2016.

Greulich, P. and Simons, B. D. (2016). Dynamic heterogeneity as a strategy of stem cell self-renewal. *Proceedings of the National Academy of Sciences*, 113(27):7509–7514.

Gudipaty, S. A. and Rosenblatt, J. (2017). Epithelial cell extrusion: Pathways and pathologies. *Seminars in Cell and Developmental Biology*, 67:132–140.

Guruharsha, K. G., Kankel, M. W., and Artavanis-Tsakonas, S. (2012). The Notch signalling system: Recent insights into the complexity of a conserved pathway. *Nature Reviews Genetics*, 13(9):654–666.

Hall, M. W., Jones, P. H., and Hall, B. A. (2019). Relating evolutionary selection and mutant clonal dynamics in normal epithelia. *Journal of the Royal Society Interface*, 16(156).

Higa, K. C. and DeGregori, J. (2019). Decoy fitness peaks, tumor suppression, and aging. *Aging Cell*, 18(3):e12938.

Hillston, J. (2005). Fluid flow approximation of pepa models. In *Second International Conference on the Quantitative Evaluation of Systems (QEST'05)*, pages 33–42.

Jain, R. B. and Wang, R. Y. (2008). Limitations of maximum likelihood estimation procedures when a majority of the observations are below the limit of detection. *Analytical Chemistry*, 80(12):4767–4772.

Johansson, F. et al. (2013). *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 0.18)*. `http://mpmath.org/`.

Jones, G. W. and Chapman, S. J. (2012). Modeling growth in biological materials. *SIAM Review*, 54(1):52–118.

Kalabis, J., Oyama, K., Okawa, T., Nakagawa, H., Michaylira, C. Z., Stairs, D. B., Figueiredo, J.-L., Mahmood, U., Diehl, J. A., Herlyn, M., and Rustgi, A. K. (2008). A subpopulation of mouse esophageal basal cells has properties of stem cells with the capacity for self-renewal and lineage specification. *Journal of Clinical Investigation*, 118(12):3860–3869.

Kansal, A., Torquato, S., IV], G. H., Chiocca, E., and Deisboeck, T. (2000). Cellular automaton of idealized brain tumor growth dynamics. *Biosystems*, 55(1):119 – 127.

Klein, A. M., Brash, D. E., Jones, P. H., and Simons, B. D. (2010). Stochastic fate of p53-mutant epidermal progenitor cells is tilted toward proliferation by UV B during preneoplasia. *Proceedings of the National Academy of Sciences of the United States of America*, 107(1):270–5.

Klein, A. M. and Simons, B. D. (2011). Universal patterns of stem cell fate in cycling adult tissues. *Development*, 138(15):3103–3111.

Kopan, R. and Ilagan, M. X. G. (2009). The canonical notch signaling pathway: Unfolding the activation mechanism. *Cell*, 137(2):216 – 233.

Le, H. Q., Ghatak, S., Yeung, C.-Y. C., Tellkamp, F., Günschmann, C., Dieterich, C., Yeroslaviz, A., Habermann, B., Pombo, A., Niessen, C. M., and Wickström, S. A. (2016). Mechanical regulation of transcription controls polycomb-mediated gene silencing during lineage commitment. *Nature Cell Biology*, 18(8):864–875.

Lee, S.-W. and Morishita, Y. (2017). Possible roles of mechanical cell elimination intrinsic to growing tissues from the perspective of tissue growth efficiency and homeostasis. *PLOS Computational Biology*, 13(7):e1005651.

Levayer, R., Dupont, C., and Moreno, E. (2016). Tissue crowding induces caspase-dependent competition for space. *Current Biology*, 26(5):670–677.

Levayer, R., Hauert, B., and Moreno, E. (2015). Cell mixing induced by myc is required for competitive tissue invasion and destruction. *Nature*, 524(7566):476–480.

Li, D. S., Zimmermann, J., and Levine, H. (2016). Modeling closure of circular wounds through coordinated collective motion. *Physical Biology*, 13(1):016006.

Lim, X., Tan, S. H., Koh, W. L. C., Chau, R. M. W., Yan, K. S., Kuo, C. J., van Amerongen, R., Klein, A. M., and Nusse, R. (2013). Interfollicular epidermal stem cells self-renew via autocrine wnt signaling. *Science*, 342(6163):1226–1230.

Lu, C. P., Polak, L., Rocha, A. S., Pasolli, H. A., Chen, S. C., Sharma, N., Blanpain, C., and Fuchs, E. (2012). Identification of stem cell populations in sweat glands and ducts reveals roles in homeostasis and wound repair. *Cell*, 150(1):136–150.

Lynch, M. D., Lynch, C. N. S., Craythorne, E., Liakath-Ali, K., Mallipeddi, R., Barker, J. N., and Watt, F. M. (2017). Spatial constraints govern competition of mutant clones in human epidermis. *Nature Communications*, 8(1):1119.

Mackenzie, I. C. (1970). Relationship between mitosis and the ordered structure of the stratum corneum in mouse epidermis. *Nature*, 226(5246):653–655.

Madan, V., Lear, J. T., and Szeimies, R.-M. (2010). Non-melanoma skin cancer. *The Lancet*, 375(9715):673 – 685.

Marchetti, L., Lombardo, R., and Priami, C. (2017). HSimulator: Hybrid stochastic/deterministic simulation of biochemical reaction networks. *Complexity*, 2017:1–12.

Marchetti, L., Priami, C., and Thanh, V. H. (2016). HRSSA – efficient hybrid stochastic simulation for spatially homogeneous biochemical reaction networks. *Journal of Computational Physics*, 317:301–317.

Marek, M. (2013). Grid anisotropy reduction for simulation of growth processes with cellular automaton. *Physica D: Nonlinear Phenomena*, 253:73–84.

Marinari, E., Mehonic, A., Curran, S., Gale, J., Duke, T., and Baum, B. (2012). Live-cell delamination counterbalances epithelial growth to limit tissue overcrowding. *Nature*, 484(7395):542–545.

Marques-Pereira, J. P. and Leblond, C. P. (1965). Mitosis and differentiation in the stratified squamous epithelium of the rat esophagus. *American Journal of Anatomy*, 117(1):73–87.

Martincorena, I., Fowler, J. C., Wabik, A., Lawson, A. R., Abascal, F., Hall, M. W., Cagan, A., Murai, K., Mahbubani, K., Stratton, M. R., Fitzgerald, R. C., Handford, P. A., Campbell, P. J., Saeb-Parsy, K., and Jones, P. H. (2018). Somatic mutant clones colonize the human esophagus with age. *Science*, 362(6417):911–917.

Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D. C., Fullam, A., Alexandrov, L. B., Tubio, J. M., Stebbings, L., Menzies, A., Widaa, S., Stratton, M. R., Jones, P. H., and Campbell, P. J. (2015). High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*, 348(6237):880–886.

Mascré, G., Dekoninck, S., Drogat, B., Youssef, K. K., Brohŕe, S., Sotiropoulou, P. a., Simons, B. D., and Blanpain, C. (2012). Distinct contribution of stem and progenitor cells to epidermal maintenance. *Nature*, 489(7415):257–62.

Massink, M., Latella, D., Bracciali, A., and Hillston, J. (2010). A combined process algebraic, agent and fluid flow approach to emergent crowd behaviour.

Melhado, R. E., Alderson, D., and Tucker, O. (2010). The changing face of esophageal cancer. *Cancers*, 2(3):1379–1404.

Mesa, K. R., Kawaguchi, K., Cockburn, K., Gonzalez, D., Boucher, J., Xin, T., Klein, A. M., and Greco, V. (2018). Homeostatic epidermal stem cell self-renewal is driven by local differentiation. *Cell Stem Cell*, 23(5):677 – 686.e4.

Metzcar, J., Wang, Y., Heiland, R., and Macklin, P. (2019). A review of cell-based computational modeling in cancer biology. *JCO Clinical Cancer Informatics*, 3(3):1–13.

Milde, F., Tauriello, G., Haberkern, H., and Koumoutsakos, P. (2014). SEM++: A particle model of cellular growth, signaling and migration. *Computational Particle Mechanics*, 1(2):211–227.

Mirams, G. R., Arthurs, C. J., Bernabeu, M. O., Bordas, R., Cooper, J., Corrias, A., Davit, Y., Dunn, S.-J., Fletcher, A. G., Harvey, D. G., Marsh, M. E., Osborne, J. M., Pathmanathan, P., Pitt-Francis, J., Southern, J., Zemzemi, N., and Gavaghan, D. J. (2013). Chaste: An open source C++ library for computational physiology and biology. *PLoS Computational Biology*, 9(3):e1002970.

Miroshnikova, Y. A., Le, H. Q., Schneider, D., Thalheim, T., Rübsam, M., Bremicker, N., Polleux, J., Kamprad, N., Tarantola, M., Wang, I., Balland, M., Niessen, C. M., Galle, J., and Wickström, S. A. (2018). Adhesion forces and cortical tension couple cell proliferation and differentiation to drive epidermal stratification. *Nature Cell Biology*, 20(1):69–80.

Moses and Olafenwa, J. (2018–). ImageAI, an open source python library built to empower developers to build applications and systems with self-contained computer vision capabilities. `https://github.com/OlafenwaMoses/ImageAI`.

Murai, K., Skrupskelyte, G., Piedrafita, G., Hall, M., Kostiou, V., Ong, S. H., Nagy, T., Cagan, A., Goulding, D., Klein, A. M., Hall, B. A., and Jones, P. H. (2018). Epidermal tissue adapts to restrain progenitors carrying clonal p53 mutations. *Cell Stem Cell*, 23(5):687–699.e8.

Newman, T. J. (2005). Modeling multicellular systems using subcellular elements. *Mathematical Biosciences and Engineering*, 2(3):613–624.

Noble, R., Burri, D., Kather, J. N., and Beerenwinkel, N. (2019). Spatial structure governs the mode of tumour evolution. *bioRxiv* 586735; 10.1101/586735.

North, M. J., Collier, N. T., Ozik, J., Tatara, E. R., Macal, C. M., Bragen, M., and Sydelko, P. (2013). Complex adaptive systems modeling with repast simphony. *Complex Adaptive Systems Modeling*, 1(3):1–26.

O'Brien, L. E. and Bilder, D. (2013). Beyond the niche: Tissue-level coordination of stem cell dynamics. *Annual Review of Cell and Developmental Biology*, 29(1):107–136.

Osborne, J. M., Fletcher, A. G., Pitt-Francis, J. M., Maini, P. K., and Gavaghan, D. J. (2017). Comparing individual-based approaches to modelling the self-organization of multicellular tissues. *PLOS Computational Biology*, 13(2):1–34.

Padgett, J. M. A. and Ilie, S. (2016). An adaptive tau-leaping method for stochastic simulations of reaction-diffusion systems. *AIP Advances*, 6(3):035217.

Page, M. E., Lombard, P., Ng, F., Göttgens, B., and Jensen, K. B. (2013). The epidermis comprises autonomous compartments maintained by distinct stem cell populations. *Cell Stem Cell*, 13(4):471–482.

Pan, Y., Heemskerk, I., Ibar, C., Shraiman, B. I., and Irvine, K. D. (2016). Differential growth triggers mechanical feedback that elevates hippo signaling. *Proceedings of the National Academy of Sciences*, 113(45):E6974–E6983.

Parigini, C. and Greulich, P. (2020). Universality of clonal dynamics poses fundamental limits to identify stem cell self-renewal strategies. *bioRxiv*.

Park, S., Greco, V., and Cockburn, K. (2016). Live imaging of stem cells: answering old questions and raising new ones. *Current Opinion in Cell Biology*, 43:30–37.

Pearson, J. W., Olver, S., and Porter, M. A. (2017). Numerical methods for the computation of the confluent and gauss hypergeometric functions. *Numerical Algorithms*, 74(3):821–866.

Pickering, C. R., Zhou, J. H., Lee, J. J., Drummond, J. A., Peng, S. A., Saade, R. E., Tsai, K. Y., Curry, J. L., Tetzlaff, M. T., Lai, S. Y., Yu, J., Muzny, D. M., Doddapaneni, H., Shinbrot, E., Covington, K. R., Zhang, J., Seth, S., Caulin, C., Clayman, G. L., El-Naggar, A. K., Gibbs, R. A., Weber, R. S., Myers, J. N., Wheeler, D. A., and Frederick, M. J. (2014). Mutational landscape of aggressive cutaneous squamous cell carcinoma. *Clinical Cancer Research*, 20(24):6582–6592.

Piedrafita, G., Kostiou, V., Wabik, A., Colom, B., Fernandez-Antoran, D., Herms, A., Murai, K., Hall, B. A., and Jones, P. H. (2020). A single-progenitor model as the unifying paradigm of epidermal and esophageal epithelial maintenance in mice. *Nature Communications*, 11(1):1429.

Potten, C. S. and Booth, C. (2002). Keratinocyte stem cells: a commentary. *Journal of Investigative Dermatology*, 119(4):888–899.

Prangle, D., Blum, M. G. B., Popovic, G., and Sisson, S. A. (2014). Diagnostic tools for approximate bayesian computation using the coverage property. *Australian & New Zealand Journal of Statistics*, 56(4):309–329.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798.

Qian, X., Tan, C., Wang, F., Yang, B., Ge, Y., Guan, Z., and Cai, J. (2016). Esophageal cancer stem cells and implications for future therapeutics. *OncoTargets and Therapy*, 9:2247–2254.

Radszuweit, M., Block, M., Hengstler, J. G., Schöll, E., and Drasdo, D. (2009). Comparing the growth kinetics of cell populations in two and three dimensions. *Physical Review E*, 79(5).

Rastrelli, M., Tropea, S., Rossi, C. R., and Alaibac, M. (2014). Melanoma: epidemiology, risk factors, pathogenesis, diagnosis and classification. *In Vivo*, 28(6):1005–1011.

Rognoni, E. and Watt, F. M. (2018). Skin cell heterogeneity in development, wound healing, and cancer. *Trends in Cell Biology*, 28(9):709–722.

Rompolas, P., Mesa, K. R., Kawaguchi, K., Park, S., Gonzalez, D., Brown, S., Boucher, J., Klein, A. M., and Greco, V. (2016). Spatiotemporal coordination of stem cell commitment during epidermal homeostasis. *Science*, 352(6292):1471–1474.

Rosekrans, S. L., Baan, B., Muncan, V., and van den Brink, G. R. (2015). Esophageal development and epithelial homeostasis. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 309(4):G216–G228.

Roshan, A., Murai, K., Fowler, J., Simons, B. D., Nikolaidou-Neokosmidou, V., and Jones, P. H. (2016). Human keratinocytes have two interconvertible modes of proliferation. *Nature Cell Biology*, 18(2):145–156.

Rossinelli, D., Bayati, B., and Koumoutsakos, P. (2008). Accelerated stochastic and hybrid methods for spatial simulations of reaction–diffusion systems. *Chemical Physics Letters*, 451(1-3):136–140.

Roy, E., Neufeld, Z., Livet, J., and Khosrotehrani, K. (2014). Concise review: Understanding clonal dynamics in homeostasis and injury through multicolor lineage tracing. *Stem Cells*, 32(12):3046–3054.

Sada, A., Jacob, F., Leung, E., Wang, S., White, B. S., Shalloway, D., and Tumbar, T. (2016). Defining the cellular lineage hierarchy in the interfollicular epidermis of adult skin. *Nature Cell Biology*, 18(6):619–631.

Sanchez-Danes, A., Hannezo, E., Larsimont, J. C., Liagre, M., Youssef, K. K., Simons, B. D., and Blanpain, C. (2016). Defining the clonal dynamics leading to mouse skin tumour initiation. *Nature*, 536(7616):298–303.

Sawas, T., Killcoyne, S., Iyer, P. G., Wang, K. K., Smyrk, T. C., Kisiel, J. B., Qin, Y., Ahlquist, D. A., Rustgi, A. K., Costa, R. J., Gerstung, M., Fitzgerald, R. C., Katzka, D. A., Noorani, A., Edwards, P. A., Grehan, N., Nutzinger, B., Hughes, C., Fidziukiewicz, E., Bornschein, J., MacRae, S., Crawte, J., Contino, G., Li, X., de la Rue, R., O'Donovan, M., Miremad, A., Malhotra, S., Tripathi, M., Tavaré, S., Lynch, A. G., Eldridge, M., Secrier, M., Bower, L., Devonshire, G., Perner, J., Jammula, S., Davies, J., Crichton, C., Carroll, N., Safranek, P., Hindmarsh, A., Sujendran, V., Hayes, S. J., Ang, Y., Preston, S. R., Oakes, S., Bagwan, I., Save, V., Skipworth, R. J., Hupp, T. R., O'Neill, J. R., Tucker, O., Beggs, A., Taniere, P., Puig, S., Underwood, T. J., Noble, F., Owsley, J., Barr, H., Shepherd, N., Old, O., Lagergren, J., Gossage, J., Davies, A., Chang, F., Zylstra, J., Mahadeva, U., Goh, V., Ciccarelli, F. D., Sanders, G., Berrisford, R., Harden, C., Bunting, D., Lewis, M., Cheong, E., Kumar, B., Parsons, S. L., Soomro, I., Kaye, P., Saunders, J., Lovat, L., Haidry, R., Eneh, V., Igali, L., Scott, M., Sothi, S., Suortamo, S., Lishman, S., Hanna, G. B., Peters, C. J., Grabowska, A., and Turkington, R. (2018). Identification of prognostic phenotypes of esophageal adenocarcinoma in 2 independent cohorts. *Gastroenterology*, 155(6):1720 – 1728.e4.

Seery, J. P. (2002). Stem cells of the oesophageal epithelium. *Journal of Cell Science*, 115(9):1783–1789.

Shiffman, D., Fry, S., and Marsh, Z. (2012). *The Nature of Code*. D. Shiffman.

Shraiman, B. I. (2005). Mechanical feedback as a possible regulator of tissue growth. *Proceedings of the National Academy of Sciences*, 102(9):3318–3323.

Sisson, S. A., Fan, Y., and Tanaka, M. M. (2007). Sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765.

Smith, C. A., Mailler, C., and Yates, C. A. (2019). Unbiased on-lattice domain growth. *Physical Review E*, 100(6):063307.

Snippert, H. J., G., L., van der Flier, Sato, T., H., J., van Es, Maaike, van den Born, Kroon-Veenboer, C., Barker, N., Klein, A. M., Jacco, van Rheenen, Simons, B. D., and Clevers, H. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing lgr5 stem cells. *Cell*, 143(1):134 – 144.

Soares, E. and Zhou, H. (2018). Master regulatory role of p63 in epidermal development and disease. *Cellular and Molecular Life Sciences*, 75(7):1179–1190.

Staddon, M. F., Bi, D., Tabatabai, A. P., Ajeti, V., Murrell, M. P., and Banerjee, S. (2018). Cooperation of dual modes of cell motility promotes epithelial stress relaxation to accelerate wound healing. *PLOS Computational Biology*, 14(10):e1006502.

Stransky, N., Egloff, A. M., Tward, A. D., Kostic, A. D., Cibulskis, K., Sivachenko, A., Kryukov, G. V., Lawrence, M. S., Sougnez, C., McKenna, A., Shefler, E., Ramos, A. H., Stojanov, P., Carter, S. L., Voet, D., Cortés, M. L., Auclair, D., Berger, M. F., Saksena, G., Guiducci, C., Onofrio, R. C., Parkin, M., Romkes, M., Weissfeld, J. L., Seethala, R. R., Wang, L., Rangel-Escareño, C., Fernandez-Lopez, J. C., Hidalgo-Miranda, A., Melendez-Zajgla, J., Winckler, W., Ardlie, K., Gabriel, S. B., Meyerson, M., Lander, E. S., Getz, G., Golub, T. R., Garraway, L. A., and Grandis, J. R. (2011). The mutational landscape of head and neck squamous cell carcinoma. *Science*, 333(6046):1157–1160.

Streichan, S. J., Hoerner, C. R., Schneidt, T., Holzer, D., and Hufnagel, L. (2014). Spatial constraints control cell proliferation in tissues. *Proceedings of the National Academy of Sciences*, 111(15):5586–5591.

Sugasawa, S. and Noma, H. (2019). A unified method for improved inference in random effects meta-analysis. *Biostatistics*, page kxz020.

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., and Dessimoz, C. (2013). Approximate bayesian computation. *PLoS Computational Biology*, 9(1):e1002803.

Tetley, R. J., Staddon, M. F., Heller, D., Hoppe, A., Banerjee, S., and Mao, Y. (2019). Tissue fluidity promotes epithelial wound healing. *Nature Physics*, 15(11):1195–1203.

The MathWorks, I. (2019). *Symbolic Math Toolbox*. Natick, Massachusetts, United States.

Toni, T. and Stumpf, M. P. H. (2009). Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics*, 26(1):104–110.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202.

Truong, A. B., Kretz, M., Ridky, T. W., Kimmel, R., and Khavari, P. A. (2006). p63 regulates proliferation and differentiation of developmentally mature keratinocytes. *Genes & Development*, 20(22):3185–3197.

Turner, B. M. and Zandt, T. V. (2012). A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology*, 56(2):69 – 85.

van der Heijden, M., Miedema, D. M., Waclaw, B., Veenstra, V. L., Lecca, M. C., Nijman, L. E., van Dijk, E., van Neerven, S. M., Lodestijn, S. C., Lenos, K. J., de Groot, N. E., Prasetyanti, P. R., Arricibita Varea, A., Winton, D. J., Medema, J. P., Morrissey, E., Ylstra, B., Nowak, M. A., Bijlsma, M. F., and Vermeulen, L. (2019). Spatiotemporal regulation of clonogenicity in colorectal cancer xenografts. *Proceedings of the National Academy of Sciences*, 116(13):6140–6145.

van Dop, W. a., Rosekrans, S. L., Uhmann, a., Jaks, V., Offerhaus, G. J. a., van den Bergh Weerman, M. a., Kasper, M., Heijmans, J., Hardwick, J. C. H., Verspaget, H. W., Hommes, D. W., Toftgard, R., Hahn, H., and van den Brink, G. R. (2012). Hedgehog signalling stimulates precursor cell accumulation and impairs epithelial maturation in the murine oesophagus. *Gut*, 62(3):348–57.

Van Liedekerke, P., Palm, M. M., Jagiella, N., and Drasdo, D. (2015). Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results. *Computational Particle Mechanics*, 2(4):401–444.

Vincent, J.-P., Fletcher, A. G., and Baena-Lopez, L. A. (2013). Mechanisms and mechanics of cell competition in epithelia. *Nature Reviews Molecular Cell Biology*, 14(9):581–591.

Wagstaff, L., Goschorska, M., Kozyrska, K., Duclos, G., Kucinski, I., Chessel, A., Hampton-O'Neil, L., Bradshaw, C. R., Allen, G. E., Rawlins, E. L., Silberzan, P., Salas, R. E. C., and Piddini, E. (2016). Mechanical cell competition kills cells via induction of lethal p53 levels. *Nature Communications*, 7(11373).

Watt, F. M., Estrach, S., and Ambler, C. A. (2008). Epidermal notch signalling: differentiation, cancer and adhesion. *Current Opinion in Cell Biology*, 20(2):171–179.

Weber, K., Thomaschewski, M., Benten, D., and Fehse, B. (2012). RGB marking with lentiviral vectors for multicolor clonal cell tracking. *Nature Protocols*, 7(5):839–849.

Wilensky, U. (1999). Netlogo. http://ccl.northwestern.edu/netlogo/, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wolfram, S. (1984). Cellular automata as models of complexity. *Nature*, 311(5985):419–424.

Yates, C. A. and Baker, R. E. (2013). Isotropic model for cluster growth on a regular lattice. *Physical Review E*, 88(2):023304.

Yokoyama, A., Kakiuchi, N., Yoshizato, T., Nannya, Y., Suzuki, H., Takeuchi, Y., Shiozawa, Y., Sato, Y., Aoki, K., Kim, S. K., Fujii, Y., Yoshida, K., Kataoka, K., Nakagawa, M. M., Inoue, Y., Hirano, T., Shiraishi, Y., Chiba, K., Tanaka, H., Sanada, M., Nishikawa, Y., Amanuma, Y., Ohashi, S., Aoyama, I., Horimatsu, T., Miyamoto, S., Tsunoda, S., Sakai, Y., Narahara, M., Brown, J. B., Sato, Y., Sawada, G., Mimori, K., Minamiguchi, S., Haga, H., Seno, H., Miyano, S., Makishima, H., Muto, M., and Ogawa, S. (2019). Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature*, 565(7739):312–317.

Publications

Published articles:

Murai, K., Skrupskelyte, G., Piedrafita, G., Hall, M., Kostiou, V., Ong, S. H., Nagy, T., Cagan, A., Goulding, D., Klein, A. M., Hall, B. A., and Jones, P. H. (2018). Epidermal tissue adapts to restrain progenitors carrying clonal p53 mutations. *Cell Stem Cell*, 23(5):687–699.e8

Piedrafita, G., Kostiou, V., Wabik, A., Colom, B., Fernandez-Antoran, D., Herms, A., Murai, K., Hall, B. A., and Jones, P. H. (2020). A single-progenitor model as the unifying paradigm of epidermal and esophageal epithelial maintenance in mice. *Nature Communications*, 11(1):1429

Cell Stem Cell

Epidermal Tissue Adapts to Restrain Progenitors Carrying Clonal *p53* Mutations

Graphical Abstract
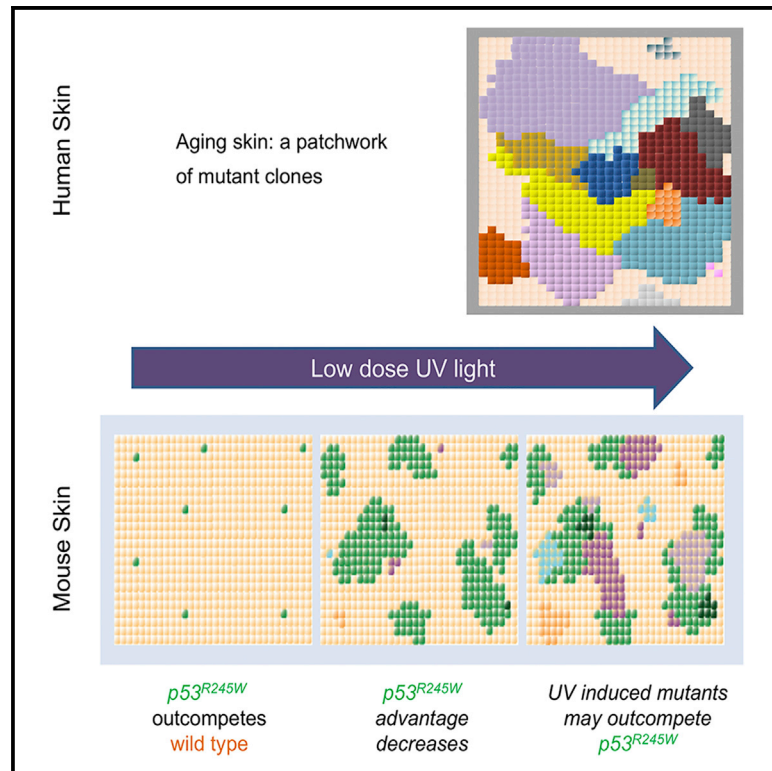
Authors

Kasumi Murai, Greta Skrupskelyte, Gabriel Piedrafita, ..., Allon M. Klein, Benjamin A. Hall, Philip H. Jones

Correspondence

pj3@sanger.ac.uk

In Brief

Normal human epidermis harbors *p53* mutant clones, very few of which form tumors. Murai et al. track *p53* mutant clones as they colonize the epidermis in a transgenic mouse model showing how the tissue adapts to the mutant cells and uncovering distinct roles for physiological UV light exposure.

Highlights

- *p53* mutant progenitors outcompete wild-type cells to colonize normal epidermis

- The epidermis adapts to the mutant clones, slowly returning toward normal

- Low-dose UV light drives *p53* mutant clone expansion in the short term

- Prolonged UV exposure generates mutant clones that may outcompete *p53* mutant cells

CellPress

CellPress

Epidermal Tissue Adapts to Restrain Progenitors Carrying Clonal *p53* Mutations

Kasumi Murai,[1,4] Greta Skrupskelyte,[1,4] Gabriel Piedrafita,[1] Michael Hall,[1,2] Vasiliki Kostiou,[2] Swee Hoe Ong,[1] Tibor Nagy,[1] Alex Cagan,[1] David Goulding,[1] Allon M. Klein,[3] Benjamin A. Hall,[2] and Philip H. Jones[1,2,5,*]
[1]Wellcome Sanger Institute, Hinxton CB10 1SA, UK
[2]MRC Cancer Unit, University of Cambridge, Hutchison-MRC Research Centre, Cambridge Biomedical Campus, Cambridge CB2 0XZ, UK
[3]Department of Systems Biology, Harvard Medical School, Harvard Medical School, Boston, MA 02115, USA
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: pj3@sanger.ac.uk
https://doi.org/10.1016/j.stem.2018.08.017

SUMMARY

Aging human tissues, such as sun-exposed epidermis, accumulate a high burden of progenitor cells that carry oncogenic mutations. However, most progenitors carrying such mutations colonize and persist in normal tissue without forming tumors. Here, we investigated tissue-level constraints on clonal progenitor behavior by inducing a single-allele *p53* mutation (*Trp53^{R245W}*; *p53^{*/wt}*), prevalent in normal human epidermis and squamous cell carcinoma, in transgenic mouse epidermis. *p53^{*/wt}* progenitors initially outcompeted wild-type cells due to enhanced proliferation, but subsequently reverted toward normal dynamics and homeostasis. Physiological doses of UV light accelerated short-term expansion of *p53^{*/wt}* clones, but their frequency decreased with protracted irradiation, possibly due to displacement by UV-induced mutant clones with higher competitive fitness. These results suggest multiple mechanisms restrain the proliferation of *p53^{*/wt}* progenitors, thereby maintaining epidermal integrity.

INTRODUCTION

Proliferating human tissues harbor substantial populations of stem/progenitor cells carrying somatic mutations linked to neoplasia and other diseases (Alexandrov et al., 2015; Jaiswal et al., 2017). This process is exemplified by the epidermis, which is exposed to UV light over decades and accumulates a high proportion of progenitors carrying UV-induced oncogenic mutations under strong evolutionary selection (Martincorena et al., 2015). For example, 4%–14% of epidermal cells in sun-exposed human skin carry non-synonymous mutations of the tumor suppressor gene *p53* (*TP53*) (Jonason et al., 1996; Martincorena et al., 2015). Despite this, aging epidermis remains histologically and functionally normal, and, in comparison with the size of the mutant progenitor population, the incidence of keratinocyte cancers is exceedingly low. Here, we investigate the mechanisms

that constrain *p53* mutant progenitors and underpin the remarkable resilience of the epidermis to mutation.

The epidermis consists of layers of keratinocytes punctuated by hair follicles and sweat ducts (Alcolea and Jones, 2014). Keratinocytes are continually shed from the tissue surface and replaced by proliferation in the basal cell layer (Figure 1A). On commitment to terminal differentiation, proliferating basal cells exit the cell cycle and migrate into the suprabasal cell layers. They then undergo a sequence of changes in gene expression and cell morphology and are ultimately shed as anucleate cornified cells. Throughout life the epidermis self-renews, matching cell production in the basal layer with cell loss from the epidermal surface (Roshan and Jones, 2012).

Various models of normal epidermal homeostasis have been proposed (Allen and Potten, 1974; Sada et al., 2016). Multiple lineage tracing and intravital imaging studies suggest the interfollicular epidermis (IFE) is maintained by a single population of progenitor cells with stochastic fate (Clayton et al., 2007; Doupé et al., 2010; Lim et al., 2013; Rompolas et al., 2016; Roshan et al., 2016). In this paradigm, progenitor cells divide to generate two progenitor daughters, two non-dividing differentiating cells or one cell of each type (Figure 1B). The outcome of individual progenitor cell divisions is unpredictable, but the probability of generating differentiated or proliferating cells is balanced. As a result, the average cell division generates one progenitor and one differentiating daughter cell across the progenitor population, achieving cellular homeostasis and ensuring the majority of clones with mutations that do not alter cell dynamics are lost by differentiation and subsequent shedding (Figure 1B) (Clayton et al., 2007).

Epidermal progenitor fate is adaptable, enabling cells to both sustain normal cell turnover and respond to injury (Lim et al., 2013; Park et al., 2017; Roshan et al., 2016). Following wounding, nearby progenitors produce more proliferating than differentiating daughters until the epidermis is repaired, after which they revert to normal, balanced behavior (Figure 1C). While the ability to modulate progenitor cell fate provides a robust mechanism of tissue repair, it also represents a potential vulnerability, as somatic mutations may drive progenitors to produce an excess of proliferating cells in the absence of injury (Alcolea et al., 2014). Unchecked, even a small bias toward proliferation may lead to clonal dominance and eventually tumor formation (Frede et al., 2016).
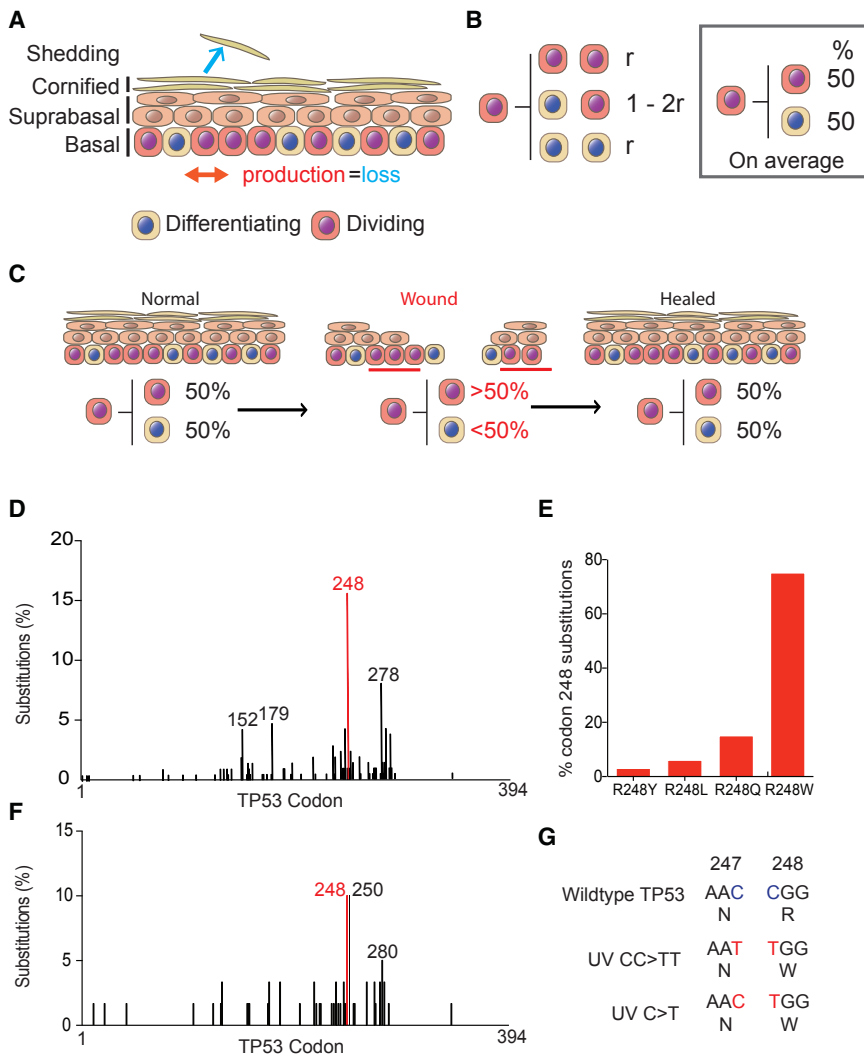
Figure 1. Cell Behavior in the Epidermis and *p53* Mutations

(A) Interfollicular epidermis (IFE). The tissue consists of layers of keratinocytes. Proliferation is confined to the basal cell layer. Differentiating basal cells exit the cell cycle and then stratify out of the basal layer, migrating through the suprabasal and cornified layers to the surface from which they are shed. In normal IFE, the rate of cell production in the basal layer (red arrow) is the same as the rate of cell loss by shedding (blue arrow).

(B) Single-progenitor model of IFE homeostasis. All dividing basal cells are functionally equivalent progenitor cells (pink). On division, a progenitor may generate two progenitors, two differentiating progeny that will cease division and stratify (beige) or one cell of each type. The outcome of a given division is unpredictable, but the likelihood (r) of producing two progenitor or two differentiating daughters is the same, so that, on average, across the population, equal proportions of progenitor and differentiating cells are generated (box).

(C) Plasticity of epidermal progenitors. Following wounding, the progenitors adjacent to the injury (red bars) switch from homeostatic behavior to producing more progenitor than differentiating progeny, until the wound is healed, and then they revert to homeostasis; numbers indicate percentages of cells generated per average cell division in each state.

(D) Distribution of TP53 missense mutations in cutaneous squamous cell carcinoma (data from COSMIC v.79, https://cancer.sanger.ac.uk/cosmic).

(E) Frequency of TP53 Codon 248 amino acid changes in cutaneous squamous cell carcinoma.

(F) Distribution of TP53 missense mutations in normal, sun-exposed human epidermis. Data from Martincorena et al., 2015.

(G) The two modes of generating TP53^{R248W} codon change from UV-signature mutations.

Here, we focus on a *p53* mutant detected in normal human epidermis, *p53*^{R248W} (Figures 1D–1G, (https://cancer.sanger.ac.uk/cosmic/) (Martincorena et al., 2015). This mutation is frequently found in human squamous cell carcinoma and has gain-of-function (GOF) properties, distinct from those of a *p53*-null allele (Muller and Vousden, 2014; Song et al., 2007). *p53*^{R248W} is a contact mutant with altered DNA-binding properties and acts in dominant-negative manner over wild-type *p53* when they are co-expressed (Song et al., 2007). Several studies from cellular systems and humanized mouse models showed that *p53*^{R248W} plays a role in the cancer phenotype through increased proliferation, drug resistance, migration, and chromosomal instability (Song et al., 2007; Zhu et al., 2015). However, recent studies suggest that some *p53* mutants including p53^{R248W} exhibit GOF attributes only in cancer cells of particular lineages and the impact of the *p53*^{R248W} mutation on normal cell behavior is unclear (Sabapathy, 2015). We hypothesized that *p53*^{R248W} may alter normal progenitor behavior to increase the likelihood of clonal persistence and the acquisition of additional mutations.

To test this hypothesis, we used a mouse model. Carcinogenesis in mouse skin has been studied extensively, and, like humans, mice may harbor mutated cells long term within normal epidermis (Berenblum and Shubik, 1949; Zhang et al., 2001). We developed a transgenic strain to track the fate of individual epidermal progenitor cells following induction of *Trp53*^{R245W}, the murine equivalent of human *TP53*^{R248W}, in a background of wild-type cells, and the impact of physiological doses of UV light on mutant cell behavior.

RESULTS

Wild-Type Progenitor Behavior in Dorsal Epidermis

We began by characterizing the behavior of *p53* wild-type (*p53*$^{wt/wt}$) progenitors in dorsal murine IFE labeled with a neutral genetic reporter by the *Ahcre*^{ERT}-inducible *cre* recombinase line. A conditional Yellow Fluorescent Protein (YFP) allele was induced at a low frequency in basal cells of *Ahcre*^{ERT} *Rosa26*$^{flYFP/wt}$ (*R-YFP*) mice (Clayton et al., 2007; Kemp et al., 2004) (Figures S1A, S2A, and S2B). Animals were sacrificed at different times and the number and location of cells in YFP-expressing (YFP$^+$) clones determined by confocal imaging (Figures S2A–S2C) (Clayton et al., 2007; Doupé et al., 2010; Page et al.,

A

Mouse genomic locus

Conditional KI allele

Conditional KI allele (after Cre recombination)

Conditional KI allele translation

B

2013). Perifollicular clones adjacent to hair follicle openings were excluded. The clonal data displayed the quantitative hallmarks of a single population of progenitors with stochastic fate (Figures S2D–S2H; STAR Methods) (Clayton et al., 2007; Doupé et al., 2010; Klein et al., 2007).

To further explore the kinetics of keratinocyte turnover, we used a transgenic cell proliferation assay based on H2BGFP dilution over a period of up to 24 weeks (Figures S1B and S2I) (Doupé et al., 2012; Sada et al., 2016). Keratinocytes showed a constant and homogeneous dilution pattern in a timescale of days. These data indicated that cycling keratinocytes divided at a similar rate (Figures S2K–S2M; STAR Methods), arguing that growth arrest and senescence, two of the stress responses mediated by *p53*, are rare in wild-type IFE (Figure S2J) (Muller and Vousden, 2014). Altogether, both lineage tracing and cell proliferation data were consistent with dorsal IFE being maintained by a single population of progenitors with balanced stochastic fate (Figures 1B and S3; STAR Methods).

Inducible *p53*$^{R245W/wt}$ Transgenic Mice

To determine whether induction of a heterozygous *p53*^{R245W} allele altered progenitor cell behavior, we developed a new conditional mouse strain, *Trp53*$^{fl-R245W-GFP/wt}$ (henceforth referred to as *p53*$^{*/wt}$). A conditional allele of *p53**, with a GFP reporter linked to the C terminus of the p53* protein by a T2A self-cleaving peptide, was targeted to the *p53* locus (Figures 2A and S4) (Trichas et al., 2008). This design allowed us to track individual *p53*$^{*/wt}$ cells in a *p53*$^{wt/wt}$ background. We confirmed

that the p53* protein had properties similar to those of the human p53^{R248W} protein and observed perturbation of p53 targets and differentiation genes in cultured keratinocytes consistent with previous reports (Freije et al., 2014; Song et al., 2007; Truong et al., 2006) (Figures S4 and S5). Following tag cleavage, a 20-amino-acid peptide remained at the C terminus of p53* protein. To test whether this peptide altered the properties of the p53* protein, we generated a second conditional mouse strain with an untagged *p53*^{R245W} mutant allele. RNA sequencing (RNA-seq) analysis of recombined primary cultures of *p53*$^{*/*}$ and the untagged *p53*$^{R245W/R245W}$ keratinocytes revealed minimal differences in transcription (Figure S5J; Table S1).

p53$^{*/wt}$ Mutant Progenitor Cells Colonize Wild-Type Epidermis

To track the fate of *p53*$^{*/wt}$ progenitors *in vivo*, we bred *Ahcre*^{ERT}*p53*$^{*/wt}$ mice and induced recombination in individual cells in adult mice (2B and Figures 3A). *p53*$^{*/wt}$-expressing clones in IFE whole mounts were imaged by confocal microscopy (Figures 3B and 3C). In contrast to control R-YFP mice, in which the area occupied by YFP$^+$ basal cells remained approximately constant after induction, the proportion of *p53*$^{*/wt}$ basal cells rose progressively indicating the mutant population had a competitive advantage over their wild-type neighbors (Figures 3D and S2D). The number of basal cells/clone was consistently higher in *p53*$^{*/wt}$ than in YFP$^+$, *p53*$^{wt/wt}$ clones at the same time point (p < 0.0001 at 1.5, 3, 6, and 12 weeks, two-tailed Mann-Whitney test, Figure 3E). By 24 weeks, the *p53*$^{*/wt}$ clones had expanded so much that they had fused, but we noted a modest but statistically significant increase in the basal cell density (cells/area, p = 0.0007 by two-tailed Mann-Whitney test) (Figures 3C, 3D, and 3F). We concluded that following induction *p53*$^{*/wt}$ cells
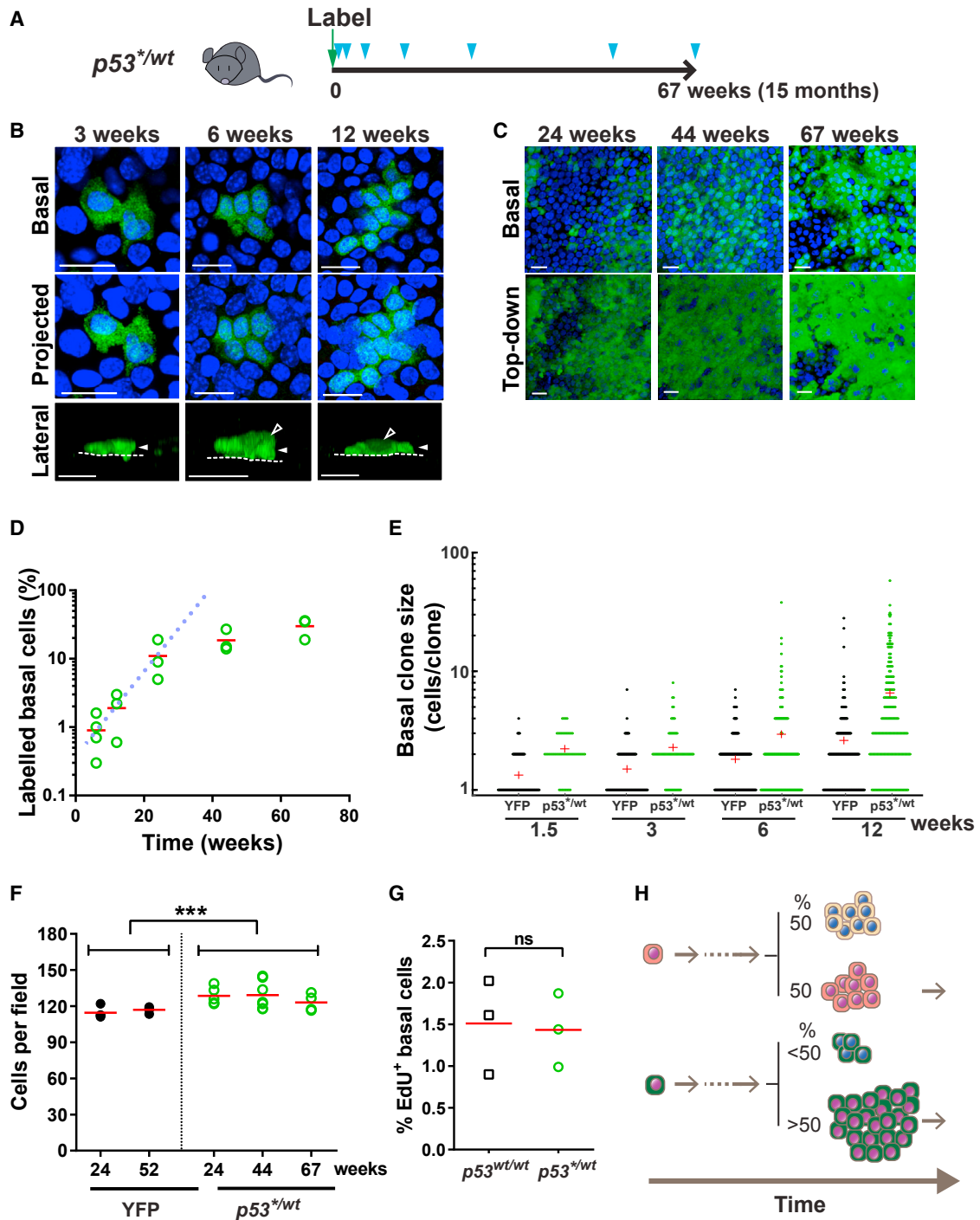
Figure 3. Heterozygous $p53^{R245W}$ ($p53^{*/wt}$) Mutant Cell Fate

(A) Protocol: clonal labeling of $Ahcre^{ERT}p53^{flR245W/wt}$ ($p53^{*/wt}$) mice followed by sampling (triangles).

(B) Rendered confocal z stacks showing typical $p53^{*/wt}$ clones in back skin epidermis. Basal, top-down view of basal layer; projected, top-down view through all nucleated cell layers; lateral, side view. Cells in basal and first suprabasal cell layers are indicated by closed and open arrowheads, respectively; dotted line, basement membrane. Green, GFP; blue, DAPI. Scale bars, 20 μm.

(C) Rendered z stacks showing basal and top-down views of typical whole mounts at times indicated. Green, GFP; blue, DAPI. Scale bars, 20 μm.

(D) Proportion of labeled basal cells at indicated time points. Averaged value from 4–5 fields per animal, n = 3 animals per time point except n = 4 animals at the 6-week time point. Red lines, mean value. Dotted line indicates the expected growth of $p53^{*/wt}$ basal cells clone area if the rate of expansion is constant.

(E) Clone size distributions (basal cells per clone). Red cross indicates mean clone size. Data are from 3–5 mice per time point. For $p53^{*/wt}$ and $RYFP$, respectively, n = 50, 93 clones at 1.5 weeks, 75, 106 clones at 3 weeks, 199, 181 clones at 6 weeks and 196, 183 clones at 12 weeks.

(legend continued on next page)

are dominant over wild-type keratinocytes, leading to colonization of the IFE.

Next, we investigated the cellular mechanism(s) that underpinned the competitive advantage of $p53^{*/wt}$ progenitors. Heightened resistance to the canonical $p53$ stress responses in the mutant cells seemed unlikely, as apoptosis was hardly detectable (0%–0.5%) and growth arrested/senescent wild-type keratinocytes were not detected in wild-type epidermis (Figure S2J; Table S2) (Clayton et al., 2007; Doupé et al., 2010). The proportion of mutant and wild-type basal cells staining for 5-ethynyl-2′-deoxyuridine (EdU), a marker of the S phase of the cell cycle, was similar, arguing that the cell-cycle time and the proportion of progenitors were not substantially altered by the mutation (Figure 3G). Within the single-progenitor paradigm, a further candidate explanation for the competitive advantage of $p53^{*/wt}$ progenitor cells was a statistical bias in the fate of mutant progenitors, leading to an excess of dividing over differentiating daughters per average cell division. Such a bias would result in an increase in the proportion of $p53^{*/wt}$ progenitors in the IFE at each successive round of cell division, even if the rate of mutant cell division was the same as that of wild-type cells (Figure 3H). Quantitative analysis showed that up to 3 months post induction the observed behavior of $p53^{*/wt}$ clones was indeed consistent with such a model, with a marked excess of divisions generating two progenitor daughters over two differentiating daughters (Figures S6A and S6B; STAR Methods).

Strikingly, beyond 24 weeks post induction, the rate of expansion of the $p53^{*/wt}$ population slowed substantially (Figure 3D). Yet, EdU data at later time points suggest that this was not due to a decrease in the cell-division rate (Table S3). This argues that the cell-fate imbalance of mutant progenitors is not cell autonomous but responds to changes in the cellular environment, such as the observed increase in basal cell density by 24 weeks (Figure 3F) (Miroshnikova et al., 2018; Roshan et al., 2016). Also, since the proportion of mutant cells adjacent to wild-type cells will decrease as the area of mutant epidermis rises, any imbalance in fate driven by competition between mutant and wild-type will inevitably decline over time. The lack of long-term clonal level data precluded further analysis to resolve between these mechanisms. However, it is clear that the reversion of $p53^{*/wt}$ cell dynamics toward normal contributed to the ability of normal tissue to tolerate mutant cells.

$p53^{*/wt}$ Perturbs Keratinocyte Differentiation and Shedding

Motivated by previous reports linking $p53$ with keratinocyte differentiation, we investigated the behavior of differentiating $p53^{*/wt}$ cells (Freije et al., 2014; Truong et al., 2006). As in wild-type IFE, mutant cell proliferation was confined to the basal layer (Figure 4A). 3D imaging of $p53^{*/wt}$ clones revealed that GFP expression was restricted to the basal and first suprabasal cell layers. This argued that mutant $p53$ was not transcribed above the lowest epidermal cell layers and/or that $p53^{*/wt}$ cells failed to complete the differentiation program and reach the upper layers of the epidermis (Figures 4B and 4C). To resolve between these possibilities, we generated $Ahcre^{ERT}p53^{*/wt}Rosa26^{Confetti/wt}$ mice carrying a conditional multicolor "Confetti" allele targeted to the Rosa26 locus (Figure 4D) (Snippert et al., 2010). Imaging of clones expressing both red fluorescent protein and GFPs revealed $p53^{*/wt}$ cells were indeed capable of terminal differentiation into cornified layer cells (Figures 4D and 4E).

Defects in keratinocyte differentiation may have a major impact on epidermal function and promote carcinogenesis, motivating us to examine the behavior of differentiating mutant cells more closely (Darido et al., 2016). Over the first 6 months after induction, the area occupied by GFP-expressing mutant suprabasal cells rose even faster than the area occupied by mutant basal cells, arguing differentiating mutant cells accumulate in the tissue compared with their wild-type counterparts (Figures 4F and 3D). In keeping with this hypothesis, the thickness of the IFE and the density of cells in the first suprabasal layer increased substantially associated with changes in cell morphology, signs of cellular disorganization and extra cell layer(s) (Figures 4G–4J and S6C).

In order to explore possible mechanisms that could lead to this scenario, we implemented computational simulations of the cell-population dynamics and tissue structure under various hypotheses incorporating parameters measured from the epidermis (STAR Methods). Briefly, we considered changes in either the proportion of symmetric divisions, stratification rate, or shedding rate along with the extent of cell-fate imbalance. Predicted changes in the proportion of $p53^{*/wt}$ cells in basal/suprabasal layers and tissue thickness were compared against experimental observations. Only $p53^{*/wt}$ progenitor cell-fate imbalance accompanied by a substantial reduction in the shedding rate was able to reproduce the patterns seen, arguing this was the most likely explanation (Figure S6D; STAR Methods). Strikingly, however, beyond 6 months, the area occupied by mutant suprabasal cells remained approximately constant while epidermal thickness decreased, arguing that mutant cell differentiation and shedding had been at least partially restored. This slow adaptation of mutant differentiating cell dynamics resulted in the preservation of tissue integrity without the development of epidermal tumors (Table S4).

The mismatch between the proportion of mutant basal and suprabasal cells has implications for cell movement within the epidermis. In normal IFE, the majority of differentiating cells stratify vertically through the suprabasal cell layers (Figure 5A) (Doupé et al., 2010; Rompolas et al., 2016). In contrast, in $p53^{*/wt}$ IFE, most mutant suprabasal cells overlie wild-type basal cells, indicating extensive lateral displacement has occurred

---

(F) Mean basal cell density in $p53^{*/wt}$ and $RYFP$ control mice. Data are means of 5 fields per mouse. n $\geq$ 4 mice at each time point/group. Exact sample numbers are in corresponding supplementary table, ***p = 0.0007 by Mann-Whitney test.

(G) Average percentage of EdU-labeled basal cells in $p53^{*/wt}$ clones at the 12-week time point compared to non-labeled cells ($p53^{wt/wt}$) in same mouse. n = 3 mice. ns, no significant difference by paired t test.

(H) Schematic illustration of $p53^{*/wt}$ cell behavior. A bias in the fate of $p53^{*/wt}$ cell can result in an increase in the proportion of $p53^{*/wt}$ progenitors in the IFE, even if the rate of mutant cell division was the same as that of wild-type cells.

See also Figure S1–S6 and Tables S2 and S7.
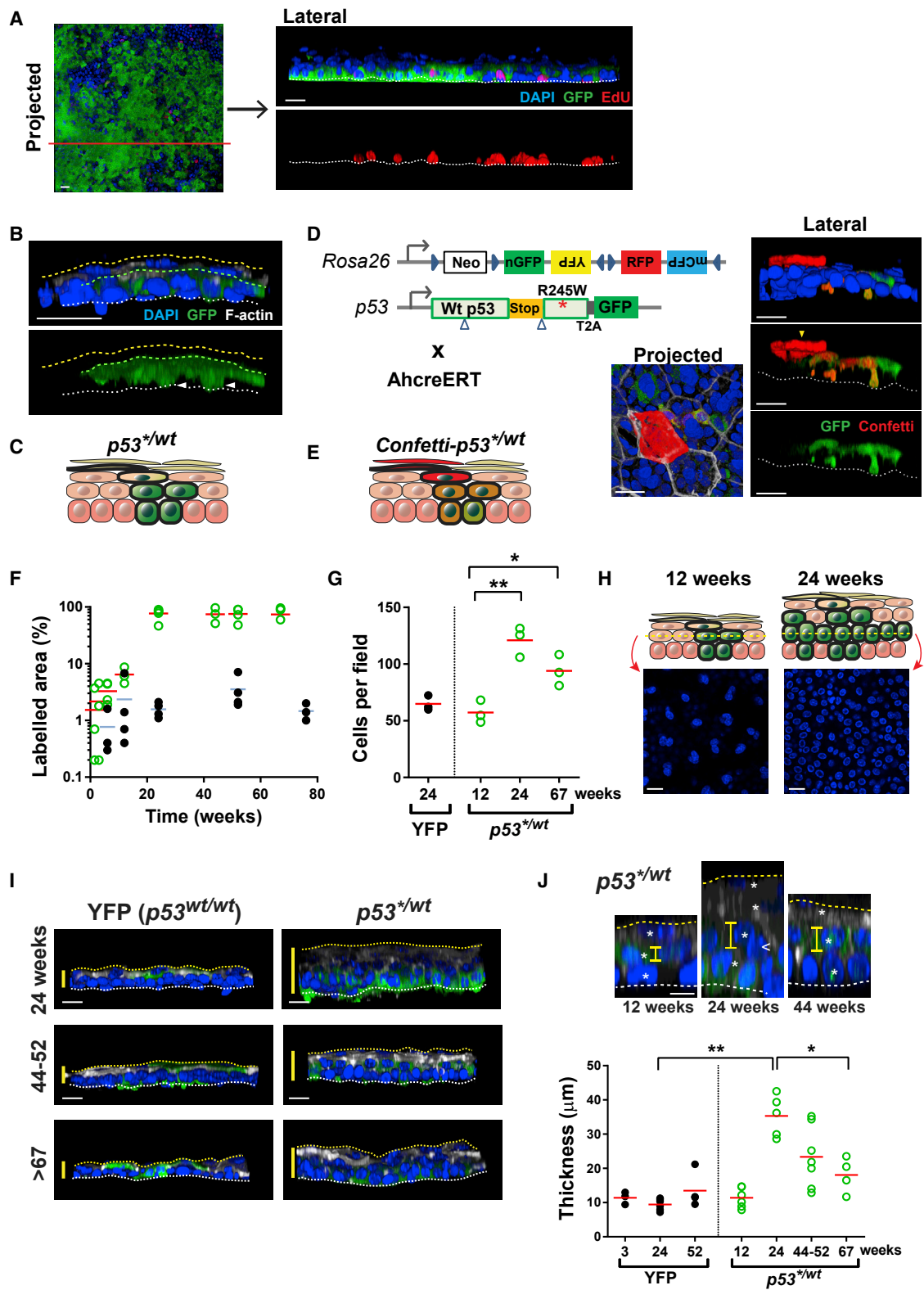
Figure 4. Behavior of Differentiating *p53*^{*/wt} Keratinocytes

(A) Rendered confocal z stacks showing *p53*^{*/wt} clone area. Lateral view shows the section of projected image (red line). White line, basement membrane; blue, DAPI; green, GFP; red, EdU. Scale bars, 20 μm.

(legend continued on next page)

(Figure 5A). We hypothesized that there may be alterations in cell-cell adhesion permissive of lateral migration in mutant IFE. Transmission electron microscopy of $p53^{*/wt}$ IFE at a year post induction revealed widespread disruption of desmosomes (Figures 5B and 5C) (Dusek and Attardi, 2011). Expression of the desmosomal proteins, DSG3 and CDH1, and levels of membrane-associated CTNNB1, which binds CDH1, were diminished in $p53^{*/wt}$ compared with adjacent wild-type areas (Figure 5D). Despite these changes, the $p53^{*/wt}$ epidermis retained its "outside-in" barrier function (Figures 5E and 5F). We concluded that the lateral migration of $p53^{*/wt}$ suprabasal cells may be facilitated by the disruption of desmosomes.

Low-Dose UV Light Exposure Drives Mutant Clone Expansion

Human epidermis is frequently exposed to low doses of UV light, below the level that causes keratinocyte apoptosis and sunburn (Jonason et al., 1996). As well as generating new mutations, repeated low doses of UV have been argued to drive the expansion of clones carrying p53 protein stabilizing mutations (Klein et al., 2010). We therefore investigated the effect of such physiological exposure.

First, we examined the impact of a short course of a low level UV-B (sub minimal erythema dose, MED, wavelengths 280–315 nm) treatment on $p53^{wt/wt}$ and $p53^{*/wt}$ progenitors in vivo. The UV dose was titrated to induce detectable DNA damage with minimal change in apoptosis (Figures S7A and S7B; Table S5). An area of shaved dorsal epidermis was exposed to daily doses of UV for 4 days per week. After 2 weeks, cre was induced and samples collected from irradiated and adjacent unexposed epidermis for up to 6 weeks (Figures 6A and 6B). In wild-type p53, R-YFP mice, the proportion of EdU$^+$ basal cells and the mean number of cells per clone were increased in irradiated compared to unexposed epidermis, consistent with UV treatment accelerating the rate of cell division (Figures 6C, S7C, and S7D). However, the area of labeled epidermis remained approximately constant despite the increased rate of epidermal turnover (Figure 6E). We concluded that the progenitor population continued to maintain the epidermis with

balanced production of progenitor and differentiating keratinocytes during low-dose UV irradiation.

In induced $p53^{*/wt}$ animals, UV treatment greatly accelerated IFE colonization by mutant cells (Figures 6B and 6F). However, the proportion of proliferating EdU$^+$ $p53^{*/wt}$ basal cells in irradiated IFE was similar to that in irradiated wild-type cells (Figures 6C and 6D). As in the unexposed epidermis, these observations indicate the fate of $p53^{*/wt}$ progenitors is biased toward proliferation. Accelerated turnover along with a fate imbalance is predicted to lead to a faster $p53^{*/wt}$ colonization (Figure 6G; STAR Methods).

In humans, low-dose UV exposure continues over decades and results in a dense patchwork of epidermal clones carrying mutations in p53 and other genes (Martincorena et al., 2015). To model this scenario, $p53^{*/wt}$ mice were induced and the skin given a single treatment with the mutagen dimethylbenzanthracene (DMBA) to load the epidermis with mutations (Abel et al., 2009). Dorsal epidermis was then treated with sub MED UV for up to 9 months, and exposed and adjacent unexposed areas compared (Figure 7A). Over the first 12 weeks the area occupied by GFP-labeled $p53^{*/wt}$ cells expanded significantly faster in the presence of UV light, as shown above (Figures 7B and 7C). However, continuing exposure resulted in a progressive decline in the $p53^{*/wt}$ population (Figures 7B–7D).

Given the short-term effect of UV treatment, the depletion of $p53^{*/wt}$ cells during long-term exposure was surprising. Several different scenarios can be envisaged to explain this observation. First, it might result from a non-cell-autonomous adaptation to a local tissue stress, e.g., such as sensitivity to increased cell density, which would also explain the restraining of $p53^{*/wt}$ expansion in non-irradiated epidermis. Alternatively, cell-autonomous factors, such as long-term UV-induced accumulation of cellular damage in p53$^{*/wt}$ cells or an adaptive increase in wild-type cell fitness may cause a decline in mutant competitiveness. Another possibility is competition between $p53^{*/wt}$ cells and mutations caused by the DMBA treatment or UV irradiation. In the last scenario, UV exposure generates new mutations with a range of possible effects on competitive fitness of mutant progenitors. Initially, $p53^{*/wt}$ cells encounter mostly wild-type cells

(B) Side view of $p53^{*/wt}$ clone at the 12-week time point. White line, basement membrane; yellow line, epidermal surface; green line, outer limit of GFP-expressing cells. Green, GFP; blue, DAPI; white, F-actin. Scale bar, 20 μm, arrowheads, basal layer cells.

(C) Schematic: typical $p53^{*/wt}$ clone with GFP expression (green) confined to basal and first suprabasal cell layers.

(D) Multicolor lineage tracing in $Ahcre^{ERT}Rosa26^{flConfetti/wt}p53^{*/wt}$ animals. Left: the multicolor confetti reporter allele encodes four fluorescent proteins. Different cre-mediated inversion and excision recombination events result in the heritable expression of one of the four fluorescent proteins depicted, nuclear GFP, yellow fluorescent protein (YFP) red fluorescent protein (RFP), or membrane cyan fluorescent protein (CFP). The fate of $p53$ mutant progenitors can be tracked in double-labeled clones even if the $p53$ locus becomes inactive in the differentiating progeny. Right: rendered z stacks of typical Confetti-$p53^{*/wt}$ double-labeled clone 12 weeks post induction of $Ahcre^{ERT}Rosa26^{Confetti}p53^{*/wt}$ mice. Images representative of 8 clones. Yellow arrowhead, cornified cell. Scale bars, 20 μm.

(E) Schematic: RFP expression reveals differentiation of $p53^{*/wt}$ clone.

(F) Projected area of labeled epidermis (%) in induced $p53^{*/wt}$ (open green circles) and $p53^{wt/wt}$ RYFP (solid black circles) mice. Data are from 5 fields per animal. n = 3–5 mice per time point/group. Red and blue lines, mean area for $p53^{*/wt}$ and $p53^{wt/wt}$ RYFP, respectively.

(G) Mean cell density in immediate suprabasal layer. Data are means of 5 fields per mouse. n = 3 mice at each time point/group except 4 fields from 1 animal are shown for YFP. *p = 0.0236, **p = 0.0035 by two-tailed unpaired t test.

(H) Single z-slice image of immediate suprabasal layer from $p53^{*/wt}$ mice at indicated time point. Schematic: typical $p53^{*/wt}$ colonized area at indicated time point. Scale bar, 20 μm.

(I) Rendered confocal z stacks showing side view of epidermal whole mounts. Yellow bars indicate the thickness of epidermis. Green, GFP; blue, DAPI; white, F-actin. Scale bar, 20 μm. White line, basement membrane; yellow line, epidermal surface.

(J) Detail analysis of change in epidermal thickness. Top, confocal images (side view) showing changes in $p53^{*/wt}$ epidermis over time. Yellow bars, length of cells in first suprabasal layer; asterisks, each cell layers. Arrowhead indicates cell in between layers. Scale bar, 20 μm. Bottom, mean thickness of epidermis at time points indicated. n ≥ 3 mice at each time point/group. *p = 0.016, **p = 0.0043 by Mann-Whitney test.

See also Figure S6 and Table S2–S4 and S7.
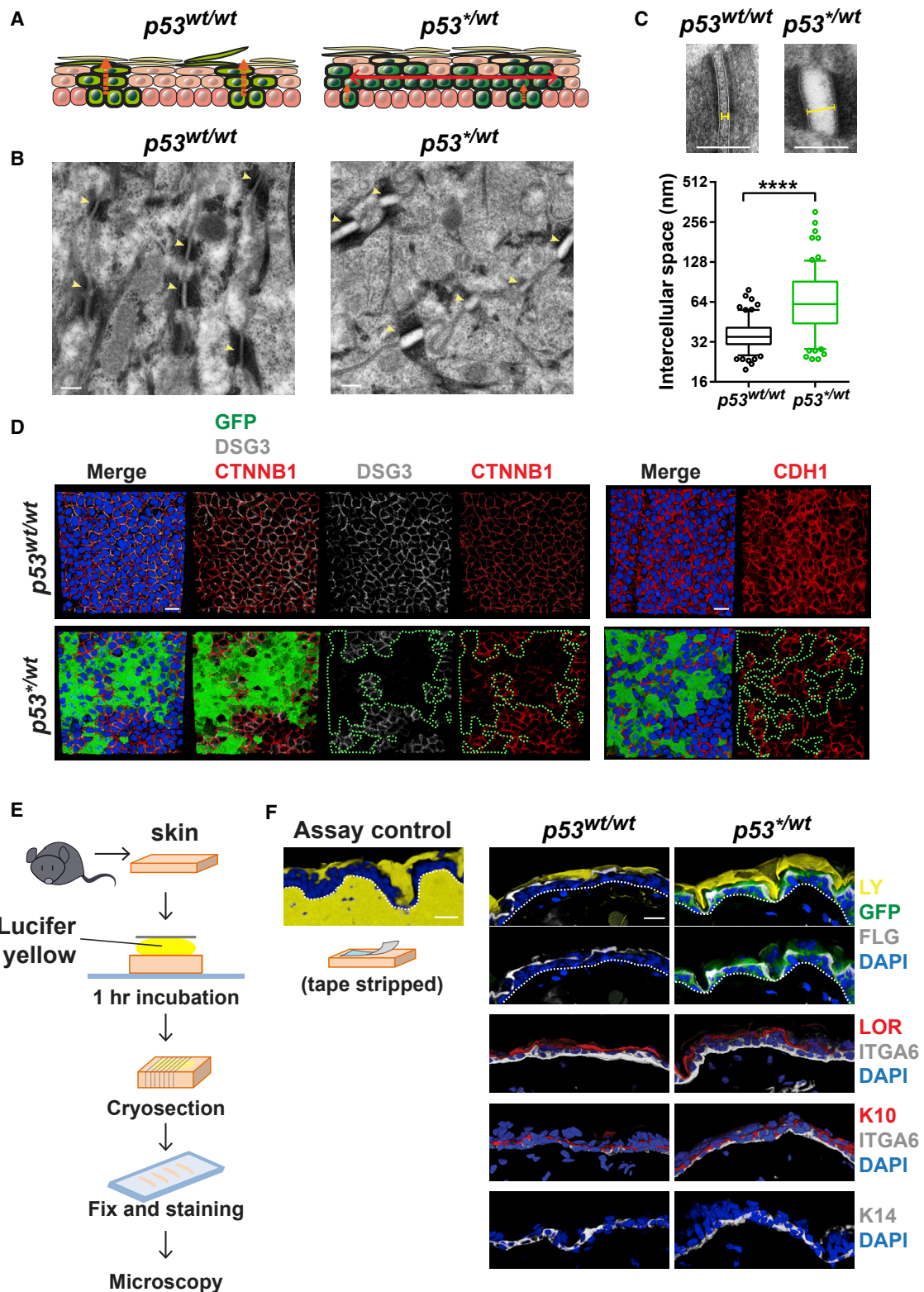
Figure 5. Disrupted Cell Adhesion in *p53*[*/wt] Epidermis at 1 Year post Induction

(A) Schematic: movement of differentiating cells is predominantly vertical in *p53*[wt/wt] mice but the excess of suprabasal over basal cells in *p53*[*/wt] IFE argues there is extensive lateral displacement of mutant cells.

(B–F) Analysis of epidermis colonized by *p53*[*/wt] at 1 year post induction. Uninduced mice were used as control. n = 3 mice per group.

(legend continued on next page)

over which they are dominant, resulting in progressive expansion of the *p53*^{*/wt} population. However, as the experiment proceeds, DMBA- or UV-induced mutant clones fitter than wild-type cells will expand and eventually collide with each other and with *p53*^{*/wt} cells. At this point, the less-fit cells would be displaced from the tissue (Figure 7E). Simple simulations embodying this hypothesis suggest such competition as a possible explanation for the observed changes in the *p53*^{*/wt} population (Figures 7F and S7E; STAR Methods; Video S1). Hence, long-term UV exposure may contribute to the decrease in the *p53*^{*/wt} population both by inducing mutations that allow cells to out-compete *p53*^{*/wt} and by providing an environment in which other mutations have an advantage over *p53*^{*/wt}.

The cell-competition hypothesis in particular makes several testable predictions. If competition is responsible for the depletion of *p53*^{*/wt} cells after 36 weeks, we would expect that clones carrying protein-altering mutations (possibly with the UV mutational signature) would be present in exposed epidermis and that the number and size of such clones would increase from 12 to 36 weeks of UV treatment. To test this, we used ultra-deep targeted exome sequencing on a panel of 74 genes linked with murine squamous carcinogenesis. Both induced and non-induced mice were included to check that the observed mutations were not merely passenger mutations within *p53*^{*/wt} clones.

While no mutations were detected in epidermis exposed to UV for 12 weeks, we identified 58 mutations in epidermis exposed to UV for 36 weeks compared with 8 in adjacent unexposed skin in 7 mice (p = 0.02, 2-tailed Mann-Whitney test) (Figures 7G and S7F; Table S6). These mutant calls correspond with very large clones as the lower limit of detection was ~0.15 mm² in area. The mutations were largely C > T nucleotide changes, consistent with the UV-light mutational signature, and the majority of mutations were non-synonymous, making it feasible for some of them to have altered cellular competitive fitness (Figures S7G and S7H). We noted recurrent cancer-associated mutations in *Trp53*, finding *R270C* mutants in UV-exposed epidermis in 4/7 mice after 36 weeks UV. The mutations found in the non-induced samples show that the expansion of new mutations does not rely on hitch-hiking within *p53*^{*/wt} clones. Taken together, these observations are consistent with clones carrying UV- or DMBA-induced mutations displacing *p53*^{*/wt} cells during prolonged UV exposure. In contrast, in the adjacent unexposed areas of the skin, the expansion of the *p53*^{*/wt} population was not impeded by competition (Figures 7B and 7C).

A further prediction of the cell-competition hypothesis is that transgenic and UV-induced mutant clones must ultimately collide. One way to visualize such events is to immuno-stain for p53. This reveals cells carrying UV-induced, protein-stabilizing mutations, such as *Trp53*^{R270C}, but does not detect the low levels of wild-type p53 or mutations that do not stabilize the protein including p53* (Figure S5H) (Jonason et al., 1996). We observed an increase in the areas covered by p53 immuno-positive cells from 12 to 36 weeks under UV exposure but barely detected any in adjacent un-irradiated epidermis at 36 weeks (Figures 7B and 7H). In several cases, these cells were in direct contact with *p53*^{*/wt} regions, infiltrating underneath differentiating edges of *p53*^{*/wt} clones (Figure 7H). We concluded that cells carrying protein-stabilizing mutations in p53 do indeed collide with *p53*^{*/wt} cells during prolonged UV exposure.

Altogether, these results suggest that, in addition to the phenotypic adaptation observed in the non-irradiated experiments, competition between mutant clones may be an important factor in constraining *p53* mutant clone expansion in the evolving mutational landscape of normal epidermis

DISCUSSION

Here, we show that the phenotype of epidermal progenitor cells carrying a heterozygous *p53* gain-of-function mutation adapts to alterations in the cellular environment. In mice protected from UV light, *p53** confers a strong competitive advantage over wild-type progenitor cells, as the average mutant cell division generates more dividing than differentiating progeny. In addition, retardation of terminal differentiation and shedding leads to mutant cell accumulation in the suprabasal layers. Unchecked, the combination of progenitor cell fate favoring proliferation and a decrease in cell shedding would be sufficient to generate tumors (Frede et al., 2016). However, the progressive attenuation of the mutant phenotype enables the epidermis to retain its functional integrity despite a high burden of *p53* mutant cells.

The mechanisms underpinning the adaptation of *p53*^{*/wt} cells remain to be determined. The slowing of expansion of the *p53*^{*/wt} basal cell population beyond 6 months post induction is accompanied by crowding of cells in the basal layer. A similar phenomenon has been observed in "imprisoned" UV-induced *p53* mutant clones and *Notch* mutant keratinocytes (Alcolea et al., 2014; Zhang et al., 2001). Keratinocyte differentiation is sensitive to mechanical forces suggesting that basal cell crowding may promote a return toward cellular homeostasis (Le et al., 2016; Roshan et al., 2016). New technologies will be required to determine the possible role of feedback effects

---

(B) Transmission electron microscopy of induced *p53*^{*/wt} and uninduced *p53*^{wt/wt} IFE. Arrows indicate desmosomes. Scale bar, 200 nm.

(C) Intercellular distance (yellow capped bar) at desmosomes in *p53*^{wt/wt} (black) and *p53*^{*/wt} (green) IFE. Box plots show median (line across box), 25th and 75th percentiles (box) and 5th and 95th percentiles (whiskers), dots are outliers (dots). n = 150 desmosomes from 3 animals per group. Scale bar, 200 nm. ****p < 0.0001 by Mann-Whitney test.

(D) Bottom-up view of rendered confocal z stacks showing the expression of cell adhesion molecules in *p53*^{*/wt} and *p53*^{wt/wt} IFE. Images are representative from 3 mice per genotype. White, DSG3; red, CTNNB1 (left panels), CDH1 (right panels); green, GFP; blue, DAPI. Green line, *p53*^{*/wt} clone area. Scale bar, 20 μm.

(E) Scheme of epidermal permeability barrier function assay.

(F) Cryosections of *p53*^{*/wt} epidermis showing normal expression of markers and barrier function compared to that of *p53*^{wt/wt}. *p53*^{*/wt} colonized over 70% at this time point and GFP (green) indicates the expression of *p53** transcript. Lucifer yellow is excluded by a normal competent epidermal surface barrier; in the assay control, barrier was removed by tape stripping. Yellow, lucifer yellow; blue, DAPI. FLG, filaggrin; LOR, loricrin; ITGA6, integrin α6; K10, keratin 10; K14, keratin 14. Scale bar, 20 μm.

See also Table S7.
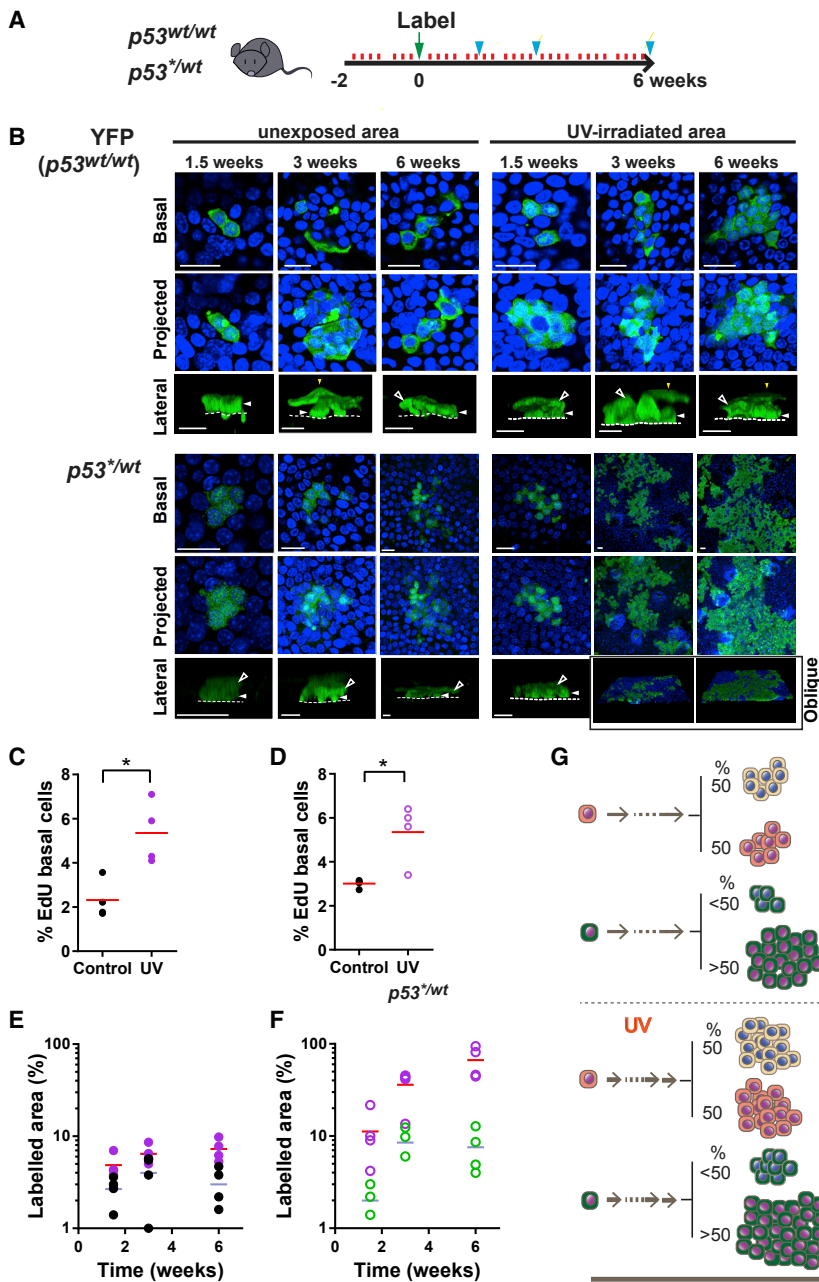
Figure 6. Short-Term Low-Dose UV Exposure Accelerates *p53*^*/wt* Colonization

(A) Protocol: *AhcreERT-p53*^*/wt* or *-RYFP* mice were exposed to a sub-erythema dose of UVB daily, 4 days per week for 2 weeks (red lines), after which labeling was induced (green arrow) and UV exposure continued; sample collection is indicated by blue arrows.

(B) Rendered confocal z stacks showing representative clones in epidermal whole mounts. Green indicates YFP in *YFP* panels and GFP in *p53*^*/wt* panels; blue, DAPI. Scale bars, 20 μm. Basal, immediate suprabasal cells and basement membrane are indicated by closed and open arrows and dotted line, respectively.

(C and D) Average percentage of EdU-labeled basal cells at the 6-week time point in induced Ah^{creERT}-*p53*^*/wt* (D) or Ah^{creERT}-*RYFP* (C) IFE. Samples were taken from UV-irradiated (purple circles) or adjacent unexposed areas (ctl, black circles). Values are mean from 5 fields per mouse. Red line indicates mean. n = 4 mice per condition. *p < 0.05 by paired t test. Percentage EdU in UV-irradiated epidermis in (D) was quantified in *p53*^*/wt* clone area.

(E and F) Projected area of labeled UV irradiated and adjacent unirradiated IFE. Values are percentage from 6 fields. n = 4 mice per time point. (E) *RYFP* (purple, UV; black, unirradiated). (F) *p53*^*/wt* (purple, UV; green, unirradiated).

(G) Schematic illustration of the effect of UV on clone behavior. UV increases the rate of cell division in both *p53*^*wt/wt* and *p53*^*/wt* cells but accelerates IFE colonization by mutant cells as *p53*^*/wt* progenitors retain a bias in fate, generating more progenitor than differentiating daughters.

See also Figure S7 and Tables S5 and S7.

dose UV light exposure may also be explained within this paradigm. UV irradiation increases the rate of cell division of both wild-type and mutant progenitors, but the imbalance of cell fate in mutant cells remains. The result is a substantial acceleration of epidermal colonization by *p53*^*/wt* cells.

DMBA/UV exposure leads to a complex and heterogeneous mutational landscape. After long-term UV exposure, the *p53*^*/wt* population declines but is not

from suprabasal cell crowding on epidermal proliferation and tissue turnover.

A persistent legacy of *p53*^*/wt* mutation is the loss of desmosomes and downregulation of expression of CDH1 and other adhesion proteins, consistent with the observation that the human *p53*^*R248W* mutant represses the expression of CDH1 *in vitro* (Dusek and Attardi, 2011; Wang et al., 2009). However, within the mixed population of mutant and wild-type cells the tissue does not exhibit a barrier or fragility defect.

In non-UV-exposed epidermis, the early competitive dominance of the *p53*^*/wt* mutant is explicable in terms of cell-autonomous effects on progenitors. The short-term effects of low-

entirely displaced. We hypothesize that individual clone fates would then depend on the chance acquisition of additional mutations and the competitive fitness of neighboring mutant clones. It remains to be seen whether competitive interactions promote, inhibit, or are neutral in carcinogenesis. Properties that allow a particular mutant to survive in early clonal competition could differ from those required for outgrowth of neoplastic cell population (Aktipis et al., 2013). The stepwise acquisition of further mutations may then be key for breaking these evolutionary bottlenecks (Nowell, 1976).

This study argues that the patchwork of clones carrying oncogenic mutations in sun-exposed human epidermis is

Figure 7. Long-Term UV Exposure Depletes Epidermis of *p53**/wt Cells

(A) Protocol: *Ahcre^ERT*-*p53**/wt mice were induced (green arrow), treated with a single dose of dimethylbenzanthracine (DMBA, black arrow), followed by repetitive sub-minimal erythema doses of UVB (red bars); blue arrows indicate sampling.

(B) Confocal z stacks showing projected views of UV-irradiated and an adjacent unexposed area of IFE. Blue, DAPI; green, GFP, reporting *p53** transcription; red, p53 indicating a p53 protein-stabilizing mutation. Scale bar, 20 μm.

(C) Labeled projected area of *p53**/wt IFE in UV-exposed (purple circles; red line indicates mean value) and adjacent unexposed areas (green circles; red lines indicate mean value). n = 3 at the 12-week time point and n = 5 at the 28- to 36-week time point. Comparison between different time points, within same animal (unexposed versus UV irradiate area), **p = 0.0046, *p = 0.036 by unpaired two-tailed t test, **p = 0.0071 by paired t test.

(D) Proportion of labeled basal cells in UV-irradiated IFE at indicated time points. Values are mean percentage from 5–8 fields per animal. n = 3 mice per time point. Red bars, mean value. *p = 0.042 by unpaired two-tailed t test.

(E) Hypothesis: effect of prolonged low-dose UV on *p53**/wt clonal dynamics. Following induction, in the absence of UV, *p53**/wt clones (green) expand progressively in a background of wild-type cells (beige). In UV-exposed IFE, a wide variety of different mutant clones arises, indicated by multiple colors, some of which may expand, outcompete, and displace *p53**/wt cells from the IFE.

(F) Simulation of clone competition under ongoing mutagenesis (see Video S1). A transgenic mutant (green) is induced at 1% frequency in a background of wild-type cells (yellow). Subsequently, new mutations (red cross, shades of green if in transgenic cells, other colors if in wild-type cells) are induced at random and assigned a fitness value as described in STAR Methods.

(G) The number of mutations detected per square millimeter by the Shearwater algorithm in each 16-mm² biopsy of mouse back skin: 3 induced and 2 non-induced mice at 12 weeks, 4 induced and

3 non-induced mice at 36 weeks. Solid circles, non-induced samples; open circles, induced samples. Green, non-exposed skin; purple, UV irradiated. **p < 0.01, *p = 0.02 by two-tailed Mann-Whitney test. Note that this method does not detect the induced and uninduced *p53** allele.

(H) Confocal z stacks showing *p53**/wt clone in direct contact with p53-immunopositive clone in UV-irradiated area (36 weeks post-induction). Blue, DAPI; green, GFP; red, p53. Scale bar, 20 μm. Dashed lines indicate outline of clones in top-down image and basement membrane in lateral view. Graph on the right shows proportion of p53 immunopositive area in UV-irradiated skin. Measurement of 8 fields of view per mouse are plotted individually. n = 3 mice per time point. Red lines indicate mean values. ****p < 0.0001 by two-tailed Mann-Whitney test.

See also Figure S7 and Tables S6 and S7.

shaped by both phenotypic adaptation and cell competition. Understanding how mutant progenitor clones interact is key to understanding not only epidermal physiology, but also the formulation of rational approaches to prevent malignant transformation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice strains and induction of allele
  - Primary culture and *in vitro* induction of allele
- METHOD DETAILS
  - Ultraviolet (UV) irradiation
  - Wholemount sample preparation
  - Immunofluorescence
  - Imaging

- ○ Transmission Electron Microscopy
- ○ Clonal imaging
- ○ Epidermal permeability barrier function
- ○ Immunoblotting
- ○ Mass Spectrometry
- ○ Transcriptome analysis
- ○ Ultra-deep targeted sequencing
- STATISTICS
- QUANTITATIVE ANALYSIS AND MODELING
- DATA AVAILABILITY
- CODE AVAILABILITY

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Gillespie (1976), (1977); Moran (1957); Tibor and Krapivsky, 2010.

REFERENCES

Abel, E.L., Angel, J.M., Kiguchi, K., and DiGiovanni, J. (2009). Multi-stage chemical carcinogenesis in mouse skin: Fundamentals and applications. Nat. Protoc. 4, 1350–1362.

Aktipis, C.A., Boddy, A.M., Gatenby, R.A., Brown, J.S., and Maley, C.C. (2013). Life history trade-offs in cancer evolution. Nat. Rev. Cancer 13, 883–892.

Alcolea, M.P., and Jones, P.H. (2014). Lineage analysis of epidermal stem cells. Cold Spring Harb. Perspect. Med. 4, a015206.

Alcolea, M.P., Greulich, P., Wabik, A., Frede, J., Simons, B.D., and Jones, P.H. (2014). Differentiation imbalance in single oesophageal progenitor cells causes clonal immortalization and field change. Nat. Cell Biol. 16, 615–622.

Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., and Stratton, M.R. (2015). Clock-like mutational processes in human somatic cells. Nat. Genet. 47, 1402–1407.

Allen, T.D., and Potten, C.S. (1974). Fine-structural identification and organization of the epidermal proliferative unit. J. Cell Sci. 15, 291–319.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57, 289–300.

Berenblum, I., and Shubik, P. (1949). The persistence of latent tumour cells induced in the mouse's skin by a single application of 9:10-dimethyl-1:2-benzanthracene. Br. J. Cancer 3, 384–386.

Clayton, E., Doupé, D.P., Klein, A.M., Winton, D.J., Simons, B.D., and Jones, P.H. (2007). A single type of progenitor cell maintains normal epidermis. Nature 446, 185–189.

Darido, C., Georgy, S.R., and Jane, S.M. (2016). The role of barrier genes in epidermal malignancy. Oncogene 35, 5705–5712.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Doupé, D.P., Klein, A.M., Simons, B.D., and Jones, P.H. (2010). The ordered architecture of murine ear epidermis is maintained by progenitor cells with random fate. Dev. Cell 18, 317–323.

Doupé, D.P., Alcolea, M.P., Roshan, A., Zhang, G., Klein, A.M., Simons, B.D., and Jones, P.H. (2012). A single progenitor population switches behavior to maintain and repair esophageal epithelium. Science 337, 1091–1093.

Dusek, R.L., and Attardi, L.D. (2011). Desmosomes: New perpetrators in tumour suppression. Nat. Rev. Cancer 11, 317–323.

Frede, J., Greulich, P., Nagy, T., Simons, B.D., and Jones, P.H. (2016). A single dividing cell population with imbalanced fate drives oesophageal tumour growth. Nat. Cell Biol. 18, 967–978.

Freije, A., Molinuevo, R., Ceballos, L., Cagigas, M., Alonso-Lecue, P., Rodriguez, R., Menendez, P., Aberdam, D., De Diego, E., and Gandarillas, A. (2014). Inactivation of p53 in human keratinocytes leads to squamous differentiation and shedding via replication stress and mitotic slippage. Cell Rep. 9, 1349–1360.

Gerstung, M., Papaemmanuil, E., and Campbell, P.J. (2014). Subclonal variant calling with multiple samples and prior knowledge. Bioinformatics 30, 1198–1204.

Gillespie, D.T. (1976). A general method for numerically simulating the stochastic evolution of coupled chemical reactions. J. Comput. Phys. 22, 403–434.

Gillespie, D. (1977). Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81, 2340–2361.

Jaiswal, S., Natarajan, P., Silver, A.J., Gibson, C.J., Bick, A.G., Shvartz, E., McConkey, M., Gupta, N., Gabriel, S., Ardissino, D., et al. (2017). Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. N. Engl. J. Med. 377, 111–121.

Jonason, A.S., Kunala, S., Price, G.J., Restifo, R.J., Spinelli, H.M., Persing, J.A., Leffell, D.J., Tarone, R.E., and Brash, D.E. (1996). Frequent clones of p53-mutated keratinocytes in normal human skin. Proc. Natl. Acad. Sci. USA 93, 14025–14029.

Kemp, R., Ireland, H., Clayton, E., Houghton, C., Howard, L., and Winton, D.J. (2004). Elimination of background recombination: Somatic induction of Cre by combined transcriptional regulation and hormone binding affinity. Nucleic Acids Res. 32, e92.

Klein, A.M., Doupé, D.P., Jones, P.H., and Simons, B.D. (2007). Kinetics of cell division in epidermal maintenance. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 76, 021910.

Klein, A.M., Brash, D.E., Jones, P.H., and Simons, B.D. (2010). Stochastic fate of p53-mutant epidermal progenitor cells is tilted toward proliferation by UV B during preneoplasia. Proc. Natl. Acad. Sci. USA 107, 270–275.

Le, H.Q., Ghatak, S., Yeung, C.-Y.C., Tellkamp, F., Günschmann, C., Dieterich, C., Yeroslaviz, A., Habermann, B., Pombo, A., Niessen, C.M., and Wickström, S.A. (2016). Mechanical regulation of transcription controls Polycomb-mediated gene silencing during lineage commitment. Nat. Cell Biol. 18, 864–875.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv, arXiv:1303.3997 https://arxiv.org/abs/1303.3997.

Lim, X., Tan, S.H., Koh, W.L., Chau, R.M., Yan, K.S., Kuo, C.J., van Amerongen, R., Klein, A.M., and Nusse, R. (2013). Interfollicular epidermal stem cells self-renew via autocrine Wnt signaling. Science 342, 1226–1230.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D.C., Fullam, A., Alexandrov, L.B., Tubio, J.M., et al. (2015). Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science 348, 880–886.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. 17, 122.

Miroshnikova, Y.A., Le, H.Q., Schneider, D., Thalheim, T., Rübsam, M., Bremicker, N., Polleux, J., Kamprad, N., Tarantola, M., Wang, I., et al. (2018). Adhesion forces and cortical tension couple cell proliferation and differentiation to drive epidermal stratification. Nat. Cell Biol. 20, 69–80.

Moran, P.A.P. (1957). Random processes in genetics. Proc. Camb. Philos. Soc. 54, 60–72.

Muller, P.A.J., and Vousden, K.H. (2014). Mutant p53 in cancer: New functions and therapeutic opportunities. Cancer Cell 25, 304–317.

Nowell, P.C. (1976). The clonal evolution of tumor cell populations. Science 194, 23–28.

Olive, K.P., Tuveson, D.A., Ruhe, Z.C., Yin, B., Willis, N.A., Bronson, R.T., Crowley, D., and Jacks, T. (2014). Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. Cell 119, 847–60.

Page, M.E., Lombard, P., Ng, F., Göttgens, B., and Jensen, K.B. (2013). The epidermis comprises autonomous compartments maintained by distinct stem cell populations. Cell Stem Cell 13, 471–482.

Park, S., Gonzalez, D.G., Guirao, B., Boucher, J.D., Cockburn, K., Marsh, E.D., Mesa, K.R., Brown, S., Rompolas, P., Haberman, A.M., et al. (2017). Tissue-scale coordination of cellular behaviour promotes epidermal wound repair in live mice. Nat. Cell Biol. 19, 155–163.

Rompolas, P., Mesa, K.R., Kawaguchi, K., Park, S., Gonzalez, D., Brown, S., Boucher, J., Klein, A.M., and Greco, V. (2016). Spatiotemporal coordination of stem cell commitment during epidermal homeostasis. Science 352, 1471–1474.

Roshan, A., and Jones, P.H. (2012). Act your age: Tuning cell behavior to tissue requirements in interfollicular epidermis. Semin. Cell Dev. Biol. 23, 884–889.

Roshan, A., Murai, K., Fowler, J., Simons, B.D., Nikolaidou-Neokosmidou, V., and Jones, P.H. (2016). Human keratinocytes have two interconvertible modes of proliferation. Nat. Cell Biol. 18, 145–156.

Sabapathy, K. (2015). The contrived mutant p53 oncogene—beyond loss of functions. Front. Oncol. 5, 276.

Sada, A., Jacob, F., Leung, E., Wang, S., White, B.S., Shalloway, D., and Tumbar, T. (2016). Defining the cellular lineage hierarchy in the interfollicular epidermis of adult skin. Nat. Cell Biol. 18, 619–631.

Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D., and Clevers, H. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell 143, 134–144.

Song, H., Hollstein, M., and Xu, Y. (2007). p53 gain-of-function cancer mutants induce genetic instability by inactivating ATM. Nat. Cell Biol. 9, 573–580.

Srinivas, S., Watanabe, T., Lin, C.S., William, C.M., Tanabe, Y., Jessell, T.M., and Costantini, F. (2001). Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Dev. Biol. 1, 4.

Tibor, A., and Krapivsky, P.L. (2010). Exact solution of a two-type branching process: Clone size distribution in cell division kinetics. J. Stat. Mech. 2010, P07028.

Trichas, G., Begbie, J., and Srinivas, S. (2008). Use of the viral 2A peptide for bicistronic expression in transgenic mice. BMC Biol. 6, 40.

Truong, A.B., Kretz, M., Ridky, T.W., Kimmel, R., and Khavari, P.A. (2006). p63 regulates proliferation and differentiation of developmentally mature keratinocytes. Genes Dev. 20, 3185–3197.

Wang, S.P., Wang, W.L., Chang, Y.L., Wu, C.T., Chao, Y.C., Kao, S.H., Yuan, A., Lin, C.W., Yang, S.C., Chan, W.K., et al. (2009). p53 controls cancer cell invasion by inducing the MDM2-mediated degradation of Slug. Nat. Cell Biol. 11, 694–704.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS 16, 284–287.

Zhang, W., Remenyik, E., Zelterman, D., Brash, D.E., and Wikonkal, N.M. (2001). Escaping the stem cell compartment: Sustained UVB exposure allows p53-mutant keratinocytes to colonize adjacent epidermal proliferating units without incurring additional mutations. Proc. Natl. Acad. Sci. USA 98, 13948–13953.

Zhu, J., Sammons, M.A., Donahue, G., Dou, Z., Vedadi, M., Getlik, M., Barsyte-Lovejoy, D., Al-awar, R., Katona, B.W., Shilatifard, A., et al. (2015). Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. Nature 525, 206–211.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Caspase 3 (CAS3)	Abcam	Cat# ab44976; PRID: AB_868674
Caspase 3 (CAS3)	Abcam	Cat# ab2302; PRID: AB_302962
Green fluorescent	Life technologies	Cat# A10262; PRID: AB_2534023
Cyclo Butane prymidine	CosmoBio	Cat# NMDND001; PRID: AB_1962813
p53 (CM5)	Vector Laboratories	Cat# VP-P956; PRID: AB_2335917
p53 (Pab1801)	Abcam	Cat# ab26; PRID: AB_303198
Phospho Serine 392 p53	Millipore	Cat# 04-244; PRID: AB_1587353
Phospho Serine 15 p53	Cell signaling	Cat# 9284; PRID: AB_331464
Acetyl Lysine 379 p53	Cell signaling	Cat# 2570; RRID:AB_823591
Loricrin	Covance	Cat# PRB-145P; PRID: AB_292095
Filaggrin	Covance	Cat# PRB-417P; PRID: AB_291632
Cytokeratin 14	Covance	Cat# PRB-155P; PRID: AB_292096
Cytokeration 10	Abcam	Cat# ab 76318; PRID: AB_1523465
Tubulin β2	Abcam	Cat# ab151318
Mouse double minute 2	Abcam	Cat# ab16895; PRID: AB_2143534
CD45	Biolegends	Cat# 103102; RRID:AB_312967
E-cadherin (CDH1)	Cell signaling	Cat# 3195; PRID: AB_10694492
Desmoglein 3 (DSG3)	Santa Cruz Biotechnology	Cat# sc-23912; RRID:AB_627422
Beta Catenin (CTNNB1)	Cell signaling	Cat# 9562; PRID: AB_331149
Phalloidin	Life technologies	Cat# A22287; PRID: AB_2620155
Lrig1	R&D systems	Cat# AF3688; RRID:AB_2138836
Alexa Fluor 647 anti-human/mouse CD49f	BioLegend	Cat# 313610; PRID:AB_493637
Bacterial and Virus Strains		
Adeno-cre	Vector Laboratories	Cat# 1045
Chemicals, Peptides, and Recombinant Proteins		
β-napthoflavone	MP Biomedicals	Cat# 156738
Tamoxifen	Sigma	Cat# N3633
Doxycycline	Sigma	Cat# D9891
Fish Skin gelatin	Sigma Aldrich	Cat# G7765
Bovine Serum Albumin	Merk MIllipore	Cat# 126575
Donkey serum	Sigma Aldrich	Cat# D9633
Goat serum	Sigma Aldrich	Cat# G9023
Lucifer yellow	Sigma	Cat# L0259
Keratinocyte Serum Free media	Invitrogen	Cat# 10744018
Epidermal growth factor	GIBCO	Cat# 10450-013
Bovine pituitary extract	GIBCO	Cat# 13028-014
Polybrene	Sigma	Cat# H9268
0.25% trypsin	Sigma	Cat# T4424
HEPES	GIBCO	Cat# 15630080
HBSS	GIBCO	Cat#14175-053
Eagle's Minimum Essential Medium	Lonza	Cat# LZBE06-174G
SimplyBlue Coomassie G-250	Thermo Scientific	Cat# LC6060
Protease and Phosphatase inhibitor	Thermo Scientific	Cat# 78415
10% Tris-HCl acrylamide gel	Bio-Rad	Cat# 161-155
Fibronectin	BD	Cat# 356008

(Continued on next page)

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Collagen Type I	BD	Cat# 354236
40, 6-diamidino-2-phenylindole (DAPI)	Sigma	Cat# D9542
Critical Commercial Assays		
Click-iTEdU imaging	Life technologies	Cat# C10086
Pierce TM Crosslink Magnetic IP/Co-IP kit	Thermo Scientific	Cat# 88805
RNeasy Mini kit	QIAGEN	Cat# 74106
Quantitect Reverse Transcription kit	QIAGEN	Cat# 205310
TaqMan Universal PCR Master Mix	Life technologies	Cat# 4304437
Lipofectamine 2000	Thermo Fisher	Cat# 11668-019
QuickSTART Bradford Dye reagents	BioRAD	Cat# 500-0202
Immobilon Western Chemiluminescent HRP substrate	Millipore	Cat# WBLUC0500
SuperSignal West Femto Chemiluminescent	Thermo Scientific	Cat# 34095
SimplyBlue Coomassie G-250 stain	Thermo Scientific	Cat# LC6060
PfuUltra II Fusion HS DNA polymerase	Agilent technologies	Cat# 600672
PCR purification kit	QIAGEN	Cat# 28106
Rapid DNA ligation kit	Roche	Cat# 11635379001
DNA MiniPrep kit	QIAGEN	Cat# 27106
Maxi Prep Endotoxin free kit	QIAGEN	Cat# 12362
Pierce TM Crosslink Magnetic immunoprecipitation (IP)/Co-IP kit	Thermo Scientific	Cat# 88805
QIAamp DNA Micro Kit	QIAGEN	Cat# 56304
Deposited Data		
RNaseq data: p53$^{wt/wt}$	This paper	ENA: ERS1755594, ERS1755602, ERS1755610, ERS1755618
RNaseq data: p53$^{*/wt}$	This paper	ENA: ERS1755595, ERS1755603, ERS1755611, ERS1755619
RNaseq data: p53$^{*/*}$	This paper	ENA: ERS1755596, ERS1755604, ERS1755612, ERS1755620
RNaseq data: p53$^{R245W/R245W}$ (untagged)	This paper	ENA: ERS1755597, ERS1755605, ERS1755613, ERS1755621
Ultra-deep targeted DNA sequencing data	This paper	ENA: ERP023080
Experimental Models: Cell Lines		
Mouse embryonic fibroblast	Laboratory of Ashok Venkitaraman	PMID:15607980
Human derived amphotrophic phoenix cell	ATCC	ATCC CRL-3213
Experimental Models: Organisms/Strains		
Mouse: C57BL/6J	The Jackson Laboratory	JAX: 000664
Mouse: $Trp53^{flR245W/flR245W}$	This paper	N/A
Mouse: $Ahcre^{ERT}$	Kemp et al., 2004	
Mouse: $Ahcre^{ERT}Trp53^{flR245W/wt}$	This paper	N/A
Mouse: $Ahcre^{ERT}R26^{flconfetti/wt}Trp53^{flR245W/wt}$	This paper	N/A
Mouse: $Ahcre^{ERT}Rosa26^{flYFP/wt}$	Clayton et al., 2007	N/A
Mouse: $R26^{M2rTA}TETO\text{-}GFP$	Doupé et al., 2012	N/A
Oligonucleotides		
Taqman assay *Mdm2*	Life technologies	Mm01233136_m1
Taqman assay *Tubb2b*	Life technologies	Mm01620966_s1
Recombinant DNA		
pBabe puro IRES-EGFP	addgene	Cat# 14430
pBabe puro IRES-EGFP p53	This paper	N/A

(*Continued on next page*)

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pBabe puro IRES-EGFP p53^{R245W}	This paper	N/A
pBabe puro IRES-EGFP p53^{R245W}-T2A	This paper	N/A
pBabe puro IRES-EGFP p53^{S389A}	This paper	N/A
Software and Algorithms		
LAS X	Leica	N/A
Volocity 6 and 6.3	Perkin Elmer	N/A
AxioVision	Zeiss	N/A
Tecnai User Interface	FEI	N/A
GraphPad	Prism 6	N/A
STAR 2.5.3a	Dobin et al., 2013	N/A
HTSeq framework version 0.6.1p1	Anders et al., 2015	N/A
R package: DESeq2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
R package: pheatmap	R Kolde, R package version 61'	https://cran.r-project.org/web/packages/pheatmap/index.html
R package: RColorBrewer	ColorBrewer palettes E Neuwirth, RC Brewer - R package version, 2014 - auckland.ac.nz	https://cran.r-project.org/web/packages/RColorBrewer/index.html
R package: clusterProfiler	Yu et al., 2012	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
R package: org.Mm.eg.db	Carlson M (2018). org.Mm.eg.db: Genome wide annotation for Mouse. R package version 3.6.0	http://bioconductor.org/packages/release/data/annotation/html/org.Mm.eg.db.html
BWA-MEM (v0.7.15)	Li, 2013	
Shearwater	Gerstung et al., 2014	https://bioconductor.org/packages/release/bioc/html/deepSNV.html
Ensembl Variant Effect Predictor (Version 84)	McLaren et al., 2016	N/A
MATLAB R2016b	MathWorks	N/A
Jupyter & Spyder 3.1 (Python 3)	Python Software Foundation	N/A
Other		
Leica TCS SP5 II and SP8	Leica	N/A
120kV FEI Spirit Biotwin	FEI	N/A
StepOne Plus Real-Time PCR system	Life technologies	N/A
UV irradiater UV-2	Tyler research Corporation	N/A
UV-irradiator CL-508M	Uvitec	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Request for reagent and resource sharing should be addressed to the Lead Contact, Philip H Jones (pj3@sanger.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice strains and induction of allele

All experiments were conducted according to UK government Home Office project licenses PPL22/2282 and PPL70/7543. Animals were maintained at specific and opportunistic pathogen free health status and were immune competent. No animals were involved in previous experiments and were drug naive prior to the start of experiments. Adult mice 12 weeks or more weeks in age were used for *in vivo* experiments. Animals were maintained on a *C57/Bl6* genetic background, housed in individually ventilated cages and fed on standard chow. Both male and female animals were used for experiments.

Trp53$^{flR245W-GFP/wt}$ knock-in mice were generated by TaconicArtemis GMBH, Germany. In the targeting vector, exons 5 to 11 of the wild-type *Trp53* gene were flanked by *loxP* sites and an additional transcriptional STOP cassette inserted between the *Trp53* 3'untranslated region (UTR) and the distal *loxP* site, in order to prevent transcriptional read through to downstream exons (Figure 2A). A second *Trp53* genomic region from exon 5 to exon 11, including the splice acceptor site of intron 4 was introduced 3' of the distal

LoxP site. This duplicated region included a Trp53[R245W] mutation introduced into exon 7 and a cassette including a glycine-serine-glycine flexible linker, a self-cleaving T2A peptide and the eGFP coding sequences inserted into exon 11, between the last codon of *Trp53* and the translation termination codon. Two positive selection markers were also introduced. A Neomycin resistance gene flanked by *Frt* sites was inserted 3′ of the 5′ *LoxP* site in exon 4, downstream of the predicted transcriptional initiation site. A Puromycin resistance gene flanked by F3 sites was placed downstream of the 3′UTR from the duplicated region. The targeting vector was generated using BAC clones from the *C57BL/6J RPCIB-731* BAC library and was fully sequenced. After transfection into the *C57BL/6N[Tac]* embryonic stem cell line, clones that had undergone homologous recombination were selected using puromycin and neomycin and screened by Southern blotting.

The conditional *Trp53* strain was generated by *Flp*-mediated deletion of the selection markers. Prior to *cre*-mediated recombination these animals (*p53[*/wt]*) express TRP53 protein from two wild-type alleles. Once the wild-type *Trp53* genomic region is deleted by *cre* both the *Trp53* mutant carrying the R245W mutation and the *eGFP* reporter are transcribed.

For lineage tracing of control, *Trp53* wild-type progenitors, the *Rosa26[flYFP/wt]* (*R-YFP*) mice which express yellow fluorescent protein (YFP) from the constitutively active *Rosa 26* locus were used (Srinivas et al., 2001).

To assess the differentiation capacity of *p53[*/wt]* mutant clones, homozygous *Ahcre[ERT]R26fl[Confetti]* animals were crossed onto *p53[*/*]* to create *Ahcre[ERT]R26fl[Confetti/wt]p53[*/wt]* animals (Snippert et al., 2010). Following induction this strain yields cells expressing one of 4 colors of reporter from the *Rosa26* locus and the GFP reporter of *p53* transcription. Cells doubly positive for red fluorescent protein and GFP were scored in experiments.

Each reporter line was crossed onto the *Ah[creERT]* line in which transcription from a transgenic *CYP1A1* (arylhydrocarbon receptor, *Ah*) promoter is normally tightly repressed (Kemp et al., 2004). Following treatment with the non-genotoxic xenobiotic β-napthoflavone the *Ah* promoter is induced and a *cre* recombinase- mutant estrogen receptor fusion protein (creERT) is expressed. In the presence of tamoxifen, the cre[ERT] protein enters the nucleus to mediate recombination.

For lineage tracing experiments, the relevant floxed reporter line was crossed onto the *Ah[creERT]* strain. Doubly transgenic male and female animals were induced by a single interaperitoneal (i.p) injection of 80 mg/kg β-napthoflavone and 1 mg tamoxifen at 11-16 weeks of age. GFP and YFP expressing clones were visualized by immunostaining with an anti-GFP antibody.

To estimate the rate of epidermal cell division a transgenic proliferation assay was used (Doupé et al., 2012). Mice doubly transgenic for the reverse tetracycline-controlled transactivator (rtTA-M2) targeted to the *Rosa26* locus and a HIST1H2BJ-EGFP fusion protein (HGFP) expressed from a tetracycline promoter element were treated with doxycycline (DOX, 2 mg ml^{-1} in drinking water sweetened with sucrose) for 4 weeks. DOX was then withdrawn and animals were culled at different time points to track HGFP dilution.

To determine the proportion of proliferating cells, 100 μg of 5-ethynyl-2′-deoxyuridine (EdU) (Life technologies) was injected intraperitoneally 1 h before culling animals.

Primary culture and *in vitro* induction of allele

Primary mouse keratinocytes were isolated directly from tail epidermis of 4-12 week-old male and female *C57/BL6* wild-type, *p53[*/wt]*, and *p53[*/*]* mice and cultured in Keratinocyte Serum Free media (KSFM, Life technologies) supplemented with 1μg/ml Epidermal growth factor, 45 μg/ml Bovine pituitary extract, 20 mM HEPES-NaOH pH7.2-7.5, 1% Penicillin-Streptomycin and CaCl$_2$ to give a final Ca^{2+} concentration of 0.02 mM for growth media (GM) and 0.6 mM for differentiation media (DM).

Cre-mediated recombination was carried out *in vitro* using an adenovirus carrying *Cre* recombinase (Adeno-*cre*, Vector Laboratories cat.no. 1045). When primary cultures reached 50%–60% confluency, cells were incubated with Adeno-*cre* at 2x10^6 pfu/ml in GM for 18 hours at 37°C. Virus was then removed by washing in Hank's Balanced Salt Solution (HBSS, Life technologies) three times. Cells maintained in GM for another 24 hours prior to use in experiments.

METHOD DETAILS

Ultraviolet (UV) irradiation

A UV irradiator UV-2 from Tyler Research Corporation was used for this study. During irradiation, animals were placed in a custom-made restrainer that restricted exposure to part of the dorsal skin. The dose of UVB-irradiation was titrated in wild-type C57BL/6 mice. A sub minimal erythema dose of UVB (750J/m^2) was determined from the appearance of the skin, Cyclobutane Pyrimidine staining and level of cleaved Caspase 3. *Ahcre[ERT]-R-YFP* and *Ahcre[ERT]-p53[*/wt]* mice were lightly shaved on the back with electric shaver 3 days prior to the start of irradiation course, then exposed to sub MED UVB daily, 4 times a week, and shaved once a week. The irradiance was monitored by dosimeter every day. For short term treatment, mice were induced with β-napthoflavone and tamoxifen after 2 weeks of irradiation and irradiation continued as before. Animals were culled at intervals as described in the text.

UV-irradiation of cultured cells was performed with a UV irradiator (Uvitec, CL-508M with 5 × 8 W 312 nm tubes). Prior to UV-irradiation, cells were rinsed in HBSS briefly, exposed to a dose of 25 mJ/cm^2 UVB in HBSS and then placed in growth media.

Wholemount sample preparation

Whole back skin was lightly shaved and treated with hair removal cream (Nair Tough Hair, Coarse/Dark). The skin was then cut into rectangular pieces of approximately 4 by 5 mm and incubated in PBS containing 5mM EDTA at 37°C for 2 hours. Samples

were transferred into PBS and the epidermis was carefully scraped off using curved scalpel while holding one corner of the skin with forceps. The epidermal wholemounts were fixed in 4% paraformaldehyde in PBS for 30 minutes and then stored in PBS at 4°C.

Immunofluorescence

For staining, wholemounts were blocked in staining buffer (0.5% Bovine Serum Albumin, 0.25% Fish Skin Gelatin, and 0.5% Triton X-100 in PBS with 10% goat or donkey serum according to the secondary antibody used) for 1 hour at room temperature. Samples were incubated with primary antibody in staining buffer overnight, washed in PBS containing 0.2% Tween-20 four times, incubated with fluorchrome-conjugated secondary antibody for 2 hours at room temperature and washed as before After the final wash, samples were incubated with 40,6-diamidino-2-phenylindole (DAPI, 1 μg ml^{-1}) in PBS at least for 20 minutes and mounted on slides using Vectashield Mounting Medium with DAPI (Vector Labs).

EdU incorporation was detected with a Click-iT imaging kit (Life technologies) according to manufacturer's instructions.

Cryosections (20 μm thickness) were fixed in 4% paraformaldehyde in PBS for 10 minutes and stained as described above.

All immunofluorescence images are representative of at least 3 animals.

Imaging

Confocal images were acquired on Leica TCS SP5 II or SP8 microscopes using 10x, 20x or 40x objectives. Typical settings for acquisition of z stacks were optimal pinhole, line average 4 scan speed 400 Hz and a resolution of 1024 × 1024 pixels or 2048 × 2048 pixels. Image analysis was performed using Volocity 6 or 6.3 image processing software (Perkin Elmer).

Transmission Electron Microscopy

Small pieces of skin (1 mm^3) were fixed at 20°C for 2 hours in 2% paraformaldehyde with 2.5% glutalaldehyde in 0.1 M sodium cacodylate buffer at pH 7.42 with 0.1% MgCl$_2$ and 0.05% CaCl$_2$. They were then rinsed three times for 10 minutes each in sodium cacodylate buffer with chlorides and placed into 1% osmium tetroxide in sodium cacodylate buffer only, at room temperature for a further 2 hours, rinsed 3 times again, mordanted with 1% tannic acid for 30 minutes and rinsed with 1% sodium sulfate for 10 minutes. The samples were then dehydrated through an ethanol series and stained *en bloc* with 2% uranyl acetate for 1 hour at the 30% ethanol stage and embedded in Epon resin. Ultrathin transverse sections were cut on a Leica UC6 microtome and mounted onto grids before contrasting with uranyl acetate and lead citrate. Finally, images were recorded on an 120kV FEI Spirit Biotwin using an F4.15 Teitz CCD camera and measurements of the junction gaps made directly using Tecnai User Interface software.

Clonal imaging

After immunostaining wholemounts, clones were imaged by confocal microscopy and the number of basal and suprabasal cells in each clone counted in live acquisition mode.

Epidermal permeability barrier function

Epidermal barrier function was assessed by Lucifer yellow (λ_{ex} 428 nm, λ_{em} 540 nm) dye diffusion assay. Mouse back skin was lightly shaved and small pieces were placed dermis side down on PBS-soaked Whatman filter paper. 10 μL of 1 mM Lucifer Yellow (Sigma L0259) in PBS was applied onto the surface of the skin and parafilm was laid on it to ensure the sample was covered with the solution. Following incubation at room temperature for 1 hour, samples were frozen in Tissue-Tek O.C.T compound (Sakura) and cyrosections (20 μm) were analyzed by confocal microscopy.

Immunoblotting

Cells were lysed in buffer containing 20 mM HEPES NaOH pH7.2-7.5, Glycerol 10%, 0.4 M NaCl, NP-40 0.5% (Sigma), 0.2 mM EDTA, 1 mM Dithiothreitol (DTT), and 0.01% Halt Protease and Phosphatase inhibitor (ThermoFisher Scientific, cat.no.78415) and centrifuged at 13000 rpm at 4°C for 10 minutes. Protein concentrations were measured using standard Bradford protein assays (BioRAD QuickSTART Bradford Dye Reagents, cat.no.500-0202). Lysates were mixed with equal amount of 2x loading buffer (100mM Tris-HCl pH6.8, 4% SDS, 20% Glycerol, Bromophenol blue and 0.2% β-mercaptoethanol) and boiled at 96°C for 5 minutes. 4-10 μg of each sample was loaded onto a 10% or 12% of SDS-polyacrylamide gel. Proteins were separated by electrophoresis and transferred onto Immobilon-P membrane (pore size 0.45 μm, Millipore). Membranes were incubated in blocking buffer (5% dried skimmed milk, PBS, 0.1% Tween-20) at room temperature for 1 hour and then with primary antibodies diluted in blocking buffer for 1 hour at room temperature or overnight at 4°C on rocking platform. After washing in 0.1% Tween-20 PBS three times, HRP conjugated secondary antibodies (Dakocytomation) diluted in 0.5% skimmed milk in PBST were applied to the membrane for 30 min at room temperature on a rocking platform followed by three washes in 0.1% Tween-20 PBS 20 min each. Proteins were detected using Immobilon Western Chemiluminescent HRP substrate (Millipore WBLUC0500) or SuperSignal West Femto Chemiluminescent Substrate (Thermo Scientific, cat.no. 34095) for high sensitivity.

Mass Spectrometry
Sample preparation

cDNAs encoding wild-type murine *p53*, *p53*^{R245W} and *p53*^{R245W} with a C-terminal cleaved T2A mutant sequence were amplified by PCR using cDNA from transgenic animals. A *p53*^{S387A} mutant was constructed by PCR mutagenesis. To produce retroviruses, pBabe

puro IRES-EGFP plasmids carrying one of the above cDNAs were transfected into Amphotrophic phoenix producer cells (ATCC) using Lipofectamine 2000, according to manufacturer's instructions (ThermoFisher). Culture media containing retrovirus was treated with polybrene (final concentration 8 μg/ml) and used to infect primary $p53^{-/-}$ mouse embryonic fibroblasts (MEFs) (Olive et al., 2004). Whole cell lysate from the MEFs was prepared as described above, desalted using Amicon Ultra-0.5 Centrifugal Filter Unit with Ultracel-10 membrane and diluted with immunoprecipitation (IP) lysis/Wash buffer.

Immunoprecipitation

Approximately 4 mg of cell lysate was used for immunoprecipitation. CM5 p53 (Vector Labs Cat No #VP-P956) antibody was cross-linked with A/G magnetic beads using Pierce TM Crosslink Magnetic IP/Co-IP kit following the manufacturer's instructions (ThermoScientific Cat No #88805). The lysate was pre-cleared and applied to 150 μl of A/G magnetic beads cross-linked with the p53 antibody. The beads were then washed and the protein was eluted by boiling at 95°C for 5 min in non-reducing buffer (2 x concentration of 100 mM Tris-HCl pH 6.8 (Sigma SLBC0806V), 4% SDS, 20% Glycerol, Bromophenol blue (BDH prolabs 101184K). The total IP product was loaded on a 10% Tris-HCl acrylamide gel (Bio-Rad Cat No #161-155) and protein bands visualized with SimplyBlue Coomassie® G-250 stain following manufacturer's instructions (ThermoScientific Cat No #LC6060). The band at appropriate size was cut out and digested into peptides for liquid chromatography tandem-mass spectrometry.

Liquid Chromatography-Mass spectrometry/Mass spectrometry (LC-MS/MS)

Experiments were performed using a nanoAcquity UPLC (Waters Corp., Milford, MA) system and an LTQ Orbitrap Velos hybrid ion trap mass spectrometer (Thermo Scientific, Waltham, MA). Separation of peptides was performed by reverse-phase chromatography using a Waters reverse-phase nano column (BEH C18, 75 μm i.d. x 250 mm, 1.7 μm particle size) at flow rate of 300 nL/min. Peptides were initially loaded onto a pre-column (Waters UPLC Trap Symmetry C18, 180 μm i.d x 20mm, 5 μm particle size) from the nanoAcquity sample manager with 0.1% formic acid for 3 minutes at a flow rate of 10 μL/min. After this period, the column valve was switched to allow the elution of peptides from the pre-column onto the analytical column. Solvent A was water + 0.1% formic acid and solvent B was acetonitrile + 0.1% formic acid. The linear gradient employed was 3%–40% B in 60 minutes.

The LC eluent was sprayed into the mass spectrometer by means of a standard Thermo Scientific nanospray source. All *m/z* values of eluting ions were measured in the Orbitrap Velos mass analyzer, set at a resolution of 30000. Data dependent scans (Top 10) were employed to automatically isolate and generate fragment ions by collision-induced dissociation in the linear ion trap, resulting in the generation of MS/MS spectra. Ions with charge states of 2+ and above were selected for fragmentation. Post-run, the data was processed using Protein Discoverer (version 2.1., ThermoFisher). Briefly, all MS/MS data were converted to mgf files and the files were then submitted to the Mascot search algorithm (Matrix Science, London UK) and searched against a Uniprot *Mus musculus* database. A fixed modification of carbamidomethyl (C) and variable modifications of oxidation (M) and deamidation (NQ) were selected. A peptide tolerance of 10 ppm (MS) and 0.6 Da (MS/MS) were also selected along with 2 missed cleavages.

Transcriptome analysis

Total RNA extraction from cultured primary mouse keratinocytes was carried out using the QIAGEN RNeasy Mini kit according to manufactures instructions (QIAGEN). cDNA was synthesized using the Quantitect Reverse Transcription kit following manufactures instructions (QIAGEN) and diluted 20-foled prior to qPCR analysis.

qRT-PCR was performed on StepOne Plus Real-Time PCR system (Life technologies). Each reaction (20μl) contained 5μl cDNA, 10 μl of TaqMan® Universal PCR Master Mix (Life technologies) and 1 μl of an appropriate taqman probe. Following PCR parameters were used: hold 95°C for 10 min, followed by 45 cycles of step 1 at 94°C for 15 s, step 2 at 60°C for 55 s and acquiring to cycling A (FAM). All reactions were performed in triplicates. Data was analyzed using The StepOnePlus Real-Time PCR System and Excel.

Target gene	TaqMan probe ID (Life technologies)
Lce3d	Mm04337274_sH
Lor	Mm01962650_s1
Mdm2	Mm01233136_m1
Cdkn1a	Mm04205640_g1
Rprm	Mm00469773_s1
Fas	Mm01204974_m1
Dusp2	Mm00839675_g1
Tubb2b	Mm00849948_g1

For RNA-seq, libraries were prepared in an automated fashion using an Agilent Bravo robot with a KAPA Standard mRNA-Seq Kit (KAPA BIOSYSTEMS). In house adaptors were ligated to 100-300 bp fragments of dsDNA. All the samples were then subject to 10 PCR cycles using sanger_168 tag set of primers and paired-end sequencing was performed on Illumina's HiSeq 2500 with 75 bp

read length. Reads were mapped using STAR 2.5.3a, the alignment files were sorted and duplicate-marked using Biobambam2 2.0.54, and the read summarization was done using the script htseq-count from version 0.6.1p1 of the HTSeq framework (Anders et al., 2015; Dobin et al., 2013). Differential expression analysis was done using the DESeq2 R package(Love et al., 2014), and the downstream pathway analysis and visualization using R (https://www.R-project.org/) and the packages Pheatmap (https://cran.r-project.org/web/packages/pheatmap/index.html), RColorBrewer (https://cran.r-project.org/web/packages/RColorBrewer/index.html), cluster-Profiler(Yu et al., 2012) and org.Mm.eg.db (http://bioconductor.org/packages/release/data/annotation/html/org.Mm.eg.db.html).

Ultra-deep targeted sequencing

Epidermal whole mounts, approximately 4 mm x 4 mm per piece, were prepared from UV-exposed and adjacent unexposed areas of the skin as described above. This protocol was run for both induced and non-induced $p53^{*/wt}$ animals. DNA was extracted using QIAGEN DNA micro kit (QIAGEN). A panel of 74 genes was chosen (see list below) based on the criteria of genes highly mutated in cutaneous squamous cell carcinomas and/or basal cell carcinomas, as well as genes frequently mutated in normal skin samples. The genes sequenced were:

Aff3, Ajuba, Arid1a, Arid2, Arid5b, Atm, Atp2a2, Bcl11b, Braf, Cacna1d, Card11, Casp8, Ccnd1, Cdkn2a, Cobll1, Crebbp, Ct*cf.*, Ctnnb1, Dclk1, Dclre1a, Dnmt3a, Ddr2, Egfr, Eif2d, Ep300, Erbb2, Erbb3, Erbb4, Ezh2, Fat1, Fat2, Fat3, Fat4, Fbxw7, Fbxo21, Fgfr3, Flt3, Grin2a, Hras, Kdm6a, Kdr, Kit, Kmt2c, Kmt2d, Kras, Lrp1b, Mtor, Nf1, Nf2, Notch1, Notch2, Notch3, Notch4, Nras, Pik3ca, Ptch1, Pten, Rb1, Ros1, Smad4, Smarca4, Smo, Sox2, Stat5b, Tert, Tet2, Tgfbr1, Tgfbr2, Trp53, Tsc1, Vhl, Zfp750, Nrf2, Keap1.

A custom bait capture panel (Agilent) was designed using Agilent SureDesign, targeting the exonic sequences of these 74 genes.

Paired-end 75bp read sequencing was performed on an Illumina HiSeq 2000_v4 machine. After removing reads for off-target capture and PCR duplicates, the average on-target coverage across samples was 1476x. The paired-end reads were aligned to the reference mouse genome (GRCm38) using BWA-MEM (v0.7.15)(Li, 2013). Variants were called using the latest version of Shearwater (unpublished). Shearwater is a variant caller designed to detect low frequency somatic variants that can be challenging to find using more conventional variant callers designed for germline variants (Gerstung et al., 2014). To avoid the probability of false positive variant calls, Shearwater builds a model of the background error rate for each base in the genome. This error model is most accurate when using matched normal samples from the same or closely related individuals to those for which variants are being called (to avoid calling common germline variants) and processed in the same way in the lab (to avoid shared artifacts introduced during sample processing and sequencing).

After running Shearwater, we detect 5888 putative variants summing across all skin biopsies. To reduce the number of false-positives we then applied a series of filtering steps. First, we removed any variants detected in both irradiated and unexposed skin biopsies from the same mouse. These are unlikely to be true somatic variants because we do not expect somatic clones to extend across two distant biopsies. To correct for multiple-hypothesis testing we then applied the Benjamini-Hochberg correction for the biopsies in each mouse independently, retaining only mutations with an adjusted q-value of less than 0.1(Benjamini and Hochberg, 1995). Finally, we removed variants without at least one supporting read from the forward and reverse strand. This resulted in a total of 67 filtered variants. The variants called here do not include the *p53** mutation, as this is removed by the Shearwater algorithm as a germline variant. The variants were annotated using Ensembl Variant Effect Predictor (Version 84)(McLaren et al., 2016)

STATISTICS

Source data and exact *P value*s for statistical tests are listed in the Table S7 for each Figure.

Statistical analysis was performed using the Graphpad Prism software. The D'Agostino-Pearson omnibus test was used to test for normality and the F-test to test for a significant difference in variance between groups. Student's unpaired t test was performed for normally distributed data where there was no significant difference in variance between groups. A two tailed paired t test was used where applicable. For non-normally distributed data, a two tailed Mann-Whitney test was performed.

No statistical method was used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments or outcome assessment.

QUANTITATIVE ANALYSIS AND MODELING

For details of quantitative analysis of wild-type (Figure S3A and S3B) and *p53* mutant progenitor cell lineage tracing data (Figures S6A and S6B), the dynamics of mutant cells in the suprabasal cell layers (Figures S6C and S6D) and a quantitative model of clonal competition during long-term ultraviolet light exposure (Figures 7F and S7E–S7G and Video S1) see STAR Methods, Quantitative Analysis.

DATA AVAILABILITY

Raw transcriptomic data can be viewed on https://www.ebi.ac.uk/ena using the following accession numbers: p53$^{wt/wt}$, ERS1755594, ERS1755602, ERS1755610, ERS1755618; p53$^{*/wt}$, ERS1755595, ERS1755603, ERS1755611, ERS1755619; p53$^{*/*}$, ERS1755596, ERS1755604, ERS1755612, ERS1755620; p53$^{R245W/R245W}$ (untag), ERS1755597, ERS1755605, ERS1755613, ERS1755621.

The accession number for the ultra-deep targeted DNA sequencing data reported in this paper is ENA: ERP023080.

CODE AVAILABILITY

Source code is accessible via Github:
https://github.com/PHJonesGroup/Murai_etal_SI_code

ARTICLE

Check for updates

A single-progenitor model as the unifying paradigm of epidermal and esophageal epithelial maintenance in mice

Gabriel Piedrafita [1,2], Vasiliki Kostiou[3], Agnieszka Wabik[1], Bartomeu Colom [1], David Fernandez-Antoran[1], Albert Herms[1], Kasumi Murai[1], Benjamin A. Hall [3✉] & Philip H. Jones[1,3✉]

In adult skin epidermis and the epithelium lining the esophagus cells are constantly shed from the tissue surface and replaced by cell division. Tracking genetically labelled cells in transgenic mice has given insight into cell behavior, but conflicting models appear consistent with the results. Here, we use an additional transgenic assay to follow cell division in mouse esophagus and the epidermis at multiple body sites. We find that proliferating cells divide at a similar rate, and place bounds on the distribution cell cycle times. By including these results in a common analytic approach, we show that data from eight lineage tracing experiments is consistent with tissue maintenance by a single population of proliferating cells. The outcome of a given cell division is unpredictable but, on average, the likelihood of producing proliferating and differentiating cells is equal, ensuring cellular homeostasis. These findings are key to understanding squamous epithelial homeostasis and carcinogenesis.

---

[1] Wellcome Sanger Institute, Hinxton CB10 1SA, UK. [2] Spanish National Cancer Research Centre (CNIO), C/Melchor Fernández Almagro 3, Madrid 29029, Spain. [3] MRC Cancer Unit, University of Cambridge, Hutchison-MRC Research Centre, Box 197, Cambridge Biomedical Campus, Cambridge CB2 0XZ, UK. ✉email: bh418@mrc-cu.cam.ac.uk; pj3@sanger.ac.uk

The squamous epithelia that cover the external surface of the body and line the mouth and esophagus consist of layers of keratinocytes. In the mouse epidermis and esophagus cell division is confined to the deepest, basal cell layer (Fig. 1a). On commitment to terminal differentiation, proliferating cells exit the cell cycle and migrate to the suprabasal cell layers, before being ultimately shed from the tissue surface. Cellular homeostasis requires that cells are generated by proliferation at the same rate at which they are shed. Further, to maintain a constant number of proliferating cells, on average each cell division must generate one daughter that will go on to divide and one that will differentiate after first exiting the cell cycle. However, the nature of the dividing cell population has been subject to controversy[1–5]. Resolving proliferating cell behavior is key for understanding not only normal tissue maintenance but also processes such as wound healing and the accumulation of somatic mutations in normal tissues during aging and carcinogenesis[6,7].

Whilst murine epidermis and esophageal epithelium share the same basic organization, there are significant differences between the tissues. The esophageal epithelium is uniform, with no appendages, while the epidermis is punctuated by hair follicles

and sweat ducts, which form distinct proliferative compartments independent of the epidermis (Fig. 1a)[5,8–10]. The structure of the epidermis also varies with body site. In typical mouse epidermis, such as that on the back (dorsum), hair follicles are frequent but there are no sweat glands[10]. In contrast, in the mouse paw epidermis hair and, particularly, sweat ducts are common in the anterior, acrosyringial region around the foot pads, while the posterior, plantar epidermis is devoid of appendages[10–12]. The ear epidermis is different again; it has uniform columns of differentiating cells, not present elsewhere[13]. Finally, the mouse tail has the most unusual structure, being a scale forming epidermis like that of chicken legs and *Crocodillia* rather than typical mammalian skin[14–16]. This structural diversity has motivated a range of studies to define the properties of proliferating cells at each site.

Genetic lineage tracing in transgenic mice has emerged as a powerful technique for tracking the behavior of cells within tissues (Fig. 1b)[17]. This is performed in mice expressing two transgenic constructs (Fig. 2a,b). The first is a genetic switch, using a bacterial recombinase enzyme *Cre*, expressed either from a transgenic promoter or targeted to a specific gene[18]. A variety of *Cre* expressing mouse strains have been used for studies of
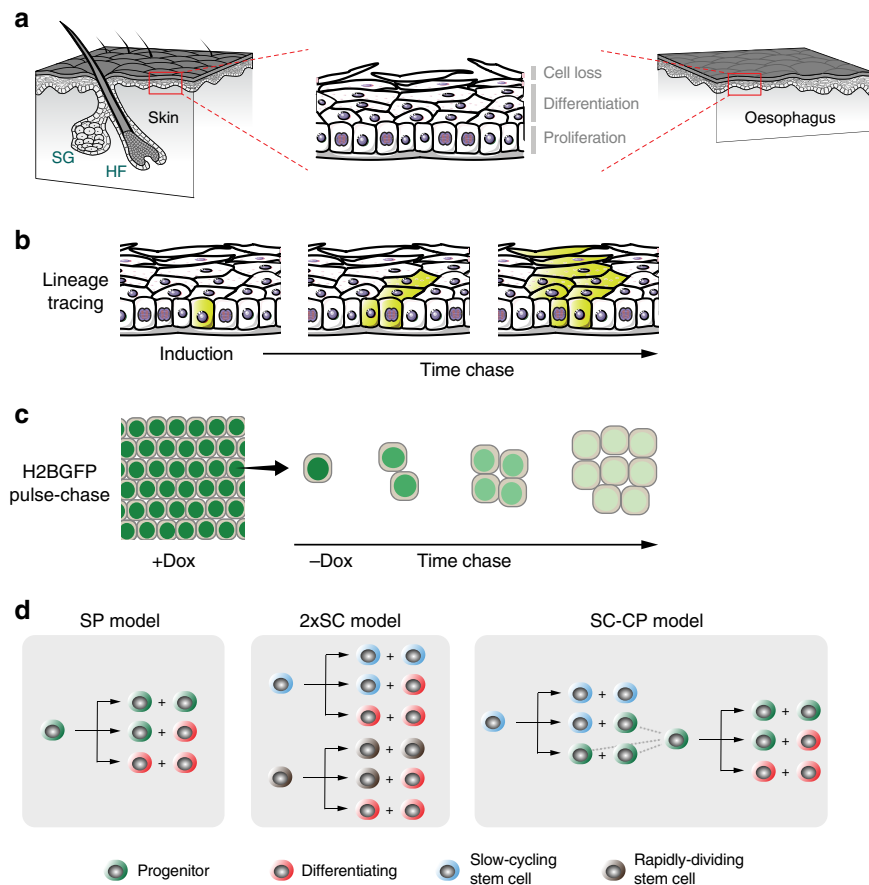
Fig. 1 Quantitative approaches to cell behavior in murine epithelium. a Structure of the stratified squamous epithelia from the interfollicular epidermis (skin) and esophagus of adult mice. Proliferating keratinocytes are located in the basal layer. Upon differentiation they migrate through suprabasal layers until they are ultimately lost by shedding. A balance should be established between cell division and cell loss to guarantee tissue homeostasis. HF, hair follicle; SG, sebaceous gland. **b** Rationale of genetic lineage tracing. Low-dose induction in transgenic mice allows recombination and conditional labeling of punctuated keratinocyte progenitors in the basal layer. These cells and their progeny remain labeled and can be tracked to study clonal dynamics over time. **c** Rationale of Histone 2B-GFP (H2BGFP) dilution experiments (a top-down view of the basal-layer plane is sketched). Transgenic-mouse keratinocytes express H2BGFP protein while on doxycycline (Dox) treatment. After Dox withdrawal cycling cells dilute their H2BGFP content with every division, allowing to study cell-proliferation rate. **d** Different stochastic cell-proliferation models invoked to explain epithelial self-renewal. Branches reflect different possible fates for a given proliferating cell upon division. SP: single-progenitor model. 2xSC: two stem-cell model, involving two independent types of proliferating cells dividing at different rates. SC-CP: stem cell-committed progenitor model, involving slow-cycling stem cells underpinning a second population of quickly-dividing progenitors.
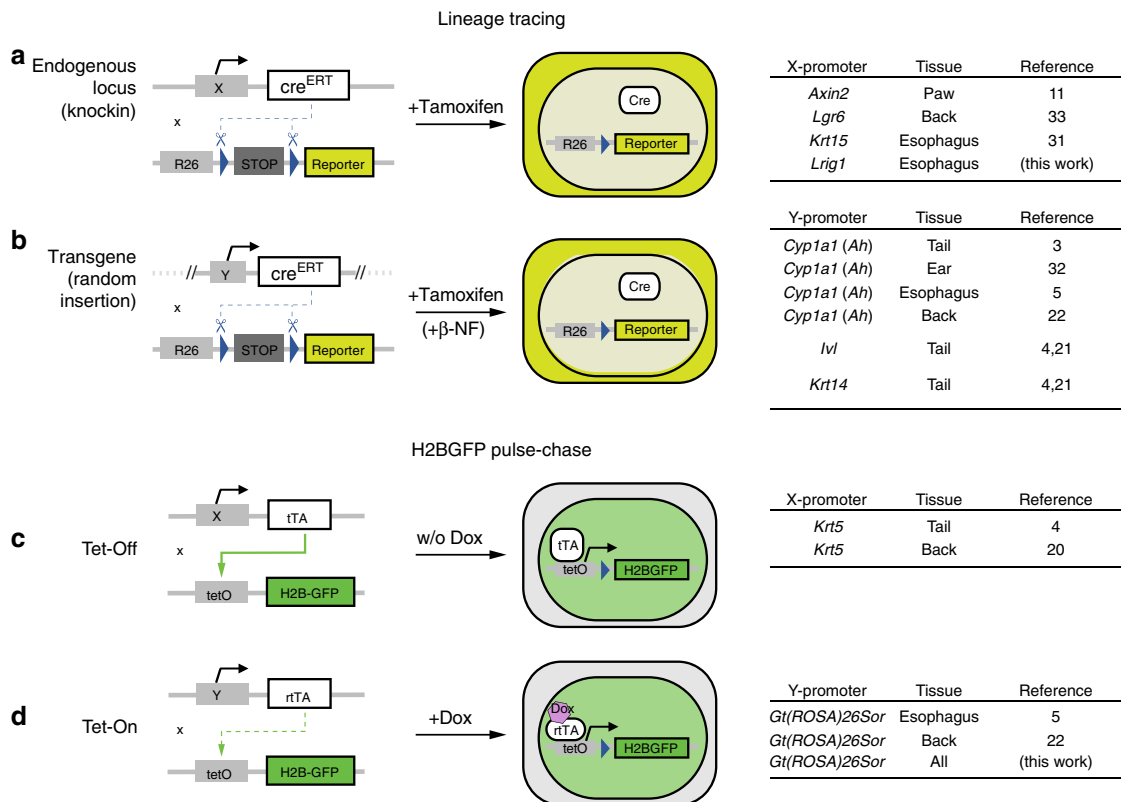
Fig. 2 Transgenic-mouse models used for lineage tracing and cell-proliferation studies. a, b Transgenic mice for lineage tracing are designed with two genetic constructs. The first codes for a bacterial *Cre* recombinase- mutant estrogen receptor fusion protein (CreERT), which can be targeted to a specific endogenous locus (**a**) or be under control of a transgenic promoter, randomly inserted in the genome (**b**). The second construct codes for a conditional fluorescent protein reporter, typically targeted to the ubiquitously expressed *Rosa26* locus. Treatment with tamoxifen induces Cre protein internalization to the nucleus, allowing expression of the reporter following Cre-mediated excision of a *loxP*-flanked STOP cassette. Specific details of constructs used in previous literature for lineage tracing in squamous epithelia are listed on the right. Note expression of *Cre* from the transgenic *Cyp1a1* arylhydrocarbon receptor, *Ah*) promoter requires additional treatment with a *Ah* inducer, β-napthoflavone (β-NF). **c, d** Transgenic mice for H2BGFP-dilution experiments are designed with a first construct, typically targeted to a constitutive promoter, coding either for a tetracycline-controlled transactivator (tTA; Tet-Off system) (**c**) or a reverse tetracycline-controlled transactivator (rtTA; Tet-On system) (**d**). A second construct codes for a Histone 2B-green fluorescent protein fusion (H2BGFP) controlled by a tetracycline-response promoter element (*tetO*; sometimes referred to as *pTRE*). Treatment with tetracycline or its derivative doxycycline (Dox) preempts tTA protein from binding to *tetO* elements in Tet-Off systems, causing repression of *pTRE*-controlled H2BGFP expression, whilst it is required for binding of rtTA to *tetO* elements in Tet-On systems, hence having an opposite effect. Dox is administered for induction and withdrawn during the H2BGFP-dilution chase in Tet-On mice, while in Tet-Off animals its application gets required for the duration of the experiment.

esophageal epithelium and epidermis (Fig. 2a,b). *Cre* is fused to a mutant hormone receptor so it is only active following treatment with a drug, giving control over when recombination is induced. Using low doses of inducing drug allows the labeling of scattered single cells. The second construct is a reporter, such as a fluorescent protein, typically targeted to the ubiquitously expressed *Gt (ROSA)26Sor* (*Rosa26*) locus. The reporter is only expressed following the excision of a "stop" cassette by *Cre* and expression persists in the progeny of the labeled cell. If the cells are labeled at a low frequency, single-cell-derived clones of reporter expressing cells result. If a representative sample of proliferating cells is labeled and their progeny tracked over a time course, statistical analysis of the evolving clone-size distributions may be used to infer cell behavior[3].

Alongside lineage tracing, a complementary transgenic assay may be used to detect cells cycling at different rates and infer the average rate of cell division (Fig. 1c). This uses a transgenic, drug regulated synthetic promoter to control expression of a protein comprising Histone 2B fused to green fluorescent protein (H2B-GFP) (Fig. 2c, d). The H2B-GFP is initially expressed at high levels in keratinocytes. Its transcription is then shut off and levels of H2B-GFP protein measured by microscopy or flow cytometry.

The stable H2B-GFP protein is diluted by cell division, so if the tissue contains cell populations dividing at different rates, the more slowly dividing cells will retain higher levels of protein[19]. Measurements of the rate of loss of fluorescence have been used to estimate the rate of cell division[4,5,20].

Lineage tracing has ruled out older deterministic models of a proliferative hierarchy of asymmetrically dividing stem cells generating 'transit amplifying' cells that undergo a fixed number of divisions prior to differentiation[3]. These models predict that clone sizes will rise and then remain stable. In multiple lineage tracing experiments, however, mean clone size has been found to increase progressively with time. However, several mutually incompatible models in which proliferating cells have stochastic fate have been proposed that do appear consistent with the data in one or more experiments (Fig. 1d; Supplementary Methods).

The simplest stochastic model, the single-progenitor (SP) hypothesis, proposes that all dividing keratinocytes are functionally equivalent and generate dividing and differentiating daughters with equal probability[3,5]. An alternative stem cell-committed progenitor (SC-CP) paradigm, applied to the epidermis proposes a hierarchy of rare, slowly cycling stem cells which generate stem and progenitor daughters. The progenitors

are biased toward differentiation so continual stem cell proliferation is required[4,21]. A third model argues that two independent populations of stem cells (2xSC) dividing at different rates exist in the epidermis[20]. These models all give comparably good fits to the results from individual experiments. However, each has been proposed on the basis of distinct data sets analyzed by different inference and fitting procedures, with limited testing of alternative hypotheses.

Motivated by the disparity of the proposed models of cell dynamics we set out to determine if a single model was consistent across multiple data sets in both esophagus and different epidermal regions. We use cell-cycle properties from the H2B-GFP dilution data to fit lineage tracing results by maximum likelihood parameter inference. We find that the data are consistent with a simple SP model of homeostasis. We also show that the fates of pairs of sister cells are anti-correlated, and that the basal layer contains a substantial proportion of cells which will differentiate rather than going on to divide.

Results

Cell-cycle times in epidermis and esophagus. Analysis of cell proliferation in epithelia offers a simple way to test the predictions of the disparate models of epithelial homeostasis by identifying the level of heterogeneity in the division rate of basal-layer cells. The SP model predicts a single-cell population dividing at the same average rate while the alternative hypotheses argue for discrete populations dividing at different average rates. We therefore investigated the dilution of H2B-GFP in the epidermis and esophagus of $R26^{M2rtTA}/TetO\text{-}H2BGFP$ mice (Fig. 3a). The animals were treated with doxycycline (Dox) for 4 weeks to induce H2BGFP expression. Dox was then withdrawn and H2BGFP protein levels in individual basal keratinocytes tracked by direct, in situ measurement of GFP fluorescence from confocal images of epithelial wholemounts at multiple time points. We examined esophagus and epidermis from plantar area of the hindpaw, ear, and tail (Figs. 3c, 4, and 5; Supplementary Movies 1–3). Optical sections through the deepest, basal cell layer were taken over at least 5 fields of view per tissue/animal and H2BGFP fluorescence quantified for all non-mitotic nuclei following image segmentation based on DAPI staining (Fig. 3b; Methods). Non-epithelial cells in the form of CD45$^+$ leukocytes, which retain high levels of H2BGFP, were excluded from the analysis, but served as internal reference for label retention[5] (e.g., Fig. 3c, insert; Figs. 4b, d, f and 5b, Supplementary Data 1). In addition, for the analysis below we included a recently published data set from dorsal epidermis performed using an identical protocol[22] (Fig. 4f; Supplementary Movie 4).

We first examined images for the presence of label-retaining cells (LRCs) (Supplementary Data 1). We found no keratinocyte LRC in the basal cell layer of the esophagus or any epidermal site other than the interscale region of the tail (Fig. 5a). Rare keratinocyte LRCs (4/1923, i.e., 0.2% of basal-layer keratinocytes) were observed in interscale epidermis, in a single animal, 18 days after DOX withdrawal (Fig. 5b). Their scarcity however suggests that they are unlikely to make a substantial contribution to tissue maintenance.

Next, we performed a quantitative analysis of the time series of the individual-cell H2BGFP intensity histograms (Supplementary Data 2). If there were multiple subpopulations of cells proliferating at different rates the distribution of H2BGFP intensities would progressively diverge, becoming wider over time. We found no evidence of such behavior in the esophagus and at multiple sites in the epidermis (Fig. 3c; Supplementary Fig. 4A, B, D, F). Specifically, several statistical tests of the modality of the distribution were applied, showing no evidence for multiple

populations (Figs. 3d, 4; Supplementary Data 3; Supplementary Methods).
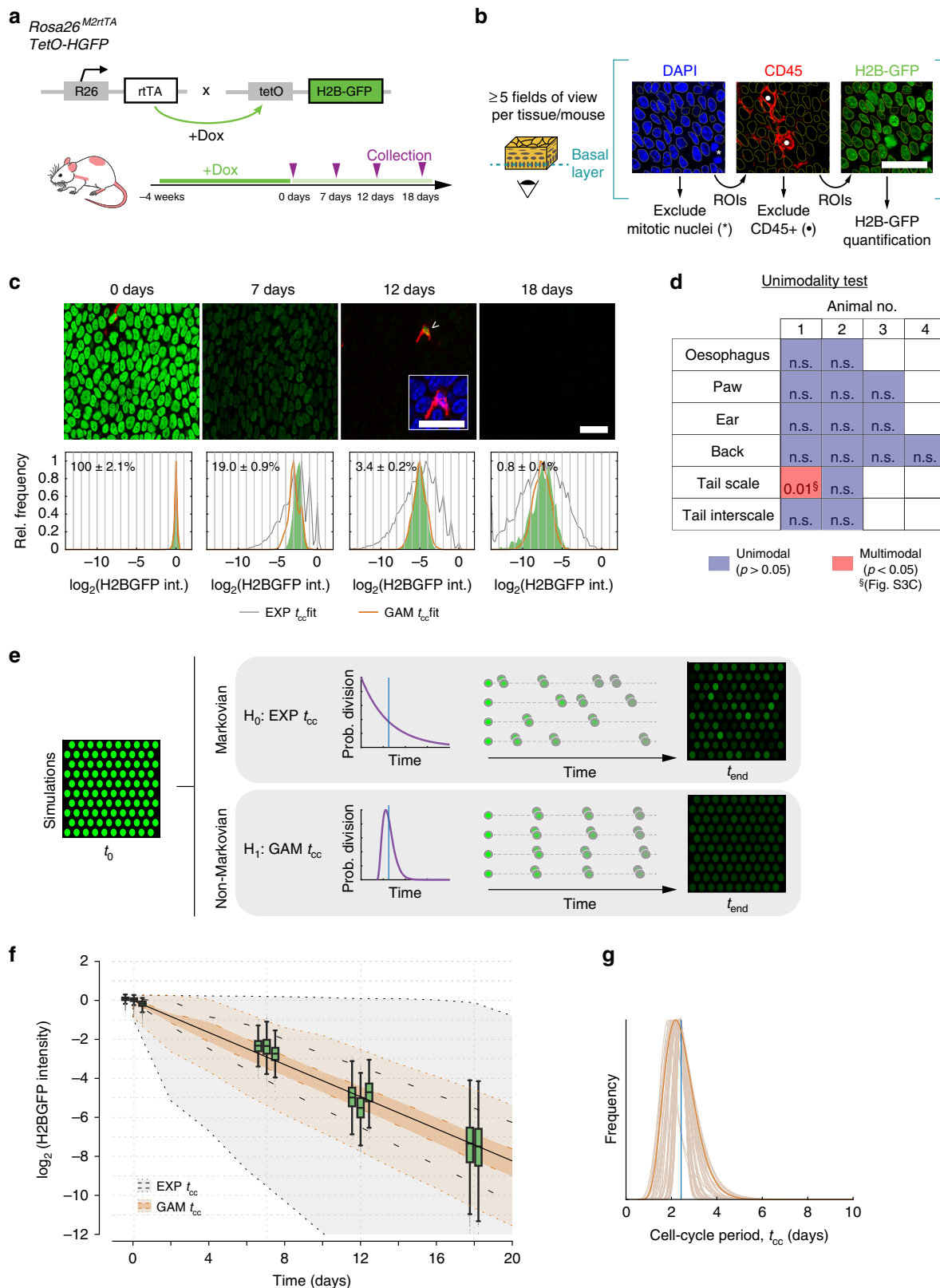
To further challenge the hypothesis that there is a single proliferating cell population, we examined whether this model can recapitulate the observed H2BGFP intensity distributions at each time point. For a given average division rate, we performed simulations of H2BGFP-dilution kinetics under a wide range of possible underlying (Gamma) distributions for individual cell-cycle times (Figs. 3e, 4b, d, f, Supplementary Methods). We find that the form of the H2BGFP histograms over time can indeed be fully described by a single population of cells, dividing within a relatively narrow range of cell-cycle times, further supporting the SP model (Figs. 3c, f, g, 4; Supplementary Data 4).

Altogether, these observations strongly argue against scenarios of heterogeneous proliferating cell populations, such as the SC-CP or 2xSC models, at all sites other than in the tail where marked variation between animals precluded reliable inference on cell-proliferation rates (Fig. 5c). We conclude that in the basal layer of the epidermis at multiple body sites and in the esophagus proliferating cells divide at a unique average rate with highly homogenous cell-cycle periods, consistent with the SP model (Table 1).

A common analytical approach to resolve cell behavior. The ability of lineage tracing to track the behavior of cohorts of proliferating cells and their progeny over time courses extending to many rounds of cell division offers the potential to validate models of homeostasis. Having established the homogeneity in the division rate of basal-layer cells, we then set out to determine whether clonal dynamics across different lineage tracing data sets were consistent with the SP paradigm.

Multiple lineage tracing studies have been published but these used distinct approaches to infer models of cell behavior and did not apply the additional constraint imposed by measuring the cell-cycle time distribution[3,4,11]. Computational simulations showed that the SP, SC-CP and 2xSC models all predict very similar development of clonal features over time, which rendered them hardly distinguishable from lineage tracing data alone (Supplementary Methods). However, as our cell-proliferation analyses do not support the SC-CP and 2xSC paradigms we focused on testing the SP model.

By incorporating the measurement of the average division rate, we could reduce the uncertainty in the parameter estimation, a problem that has been generally overlooked in these stochastic models (Supplementary Fig. 1A-B). For example, a relatively high division rate and modest proportion of symmetric division outcomes predict a similar clone-size distribution to those with a slower turnover rate but higher level of symmetric divisions. In turn, whilst long-term model predictions on clone-size distributions remained largely unaffected by the assumptions on the cell-cycle time distribution, introducing realistic estimates for the distribution of individual cell-cycle lengths affected short-term clone-size predictions, impacting on the inferred parameter values (Supplementary Fig. 2). This is due to the probability of a chain of consecutive division events deviating from the average rate, for example a run of several consecutive divisions shorter than average cell-cycle times (Supplementary Fig. 2a). This results in a broadening of the clone-size distribution at early time points after labeling. At later times, where many rounds of division have occurred in each clone, these random cycle time variations regress toward the mean cycle time of the population (Supplementary Fig. 2a). This disputes most analyses that use a Markovian implementation which makes the biologically implausible assumption that cell cycle times are distributed exponentially (i.e., the likeliest time for a cell to divide is immediately after the division

a
$Rosa26^{M2rtTA}$
$TetO$-$HGFP$

b
≥5 fields of view per tissue/mouse

Basal layer

DAPI   CD45   H2B-GFP

Exclude mitotic nuclei (*)   ROIs   Exclude CD45+ (•)   ROIs   H2B-GFP quantification

c
0 days   7 days   12 days   18 days

100 ± 2.1%   19.0 ± 0.9%   3.4 ± 0.2%   0.8 ± 0.1%

Rel. frequency
log₂(H2BGFP int.)

— EXP t_{cc}fit
— GAM t_{cc}fit

d
Unimodality test

	Animal no.			
	1	2	3	4
Oesophagus	n.s.	n.s.		
Paw	n.s.	n.s.	n.s.	
Ear	n.s.	n.s.	n.s.	
Back	n.s.	n.s.	n.s.	n.s.
Tail scale	0.01§	n.s.		
Tail interscale	n.s.	n.s.		

Unimodal ($p > 0.05$)   Multimodal ($p < 0.05$) §(Fig. S3C)

e
Simulations t_0

Markovian H_0: EXP t_{cc}

Prob. division   Time   Time t_{end}

Non-Markovian H_1: GAM t_{cc}

Prob. division   Time   Time t_{end}

f
log₂ (H2BGFP intensity)   Time (days)

EXP t_{cc}
GAM t_{cc}

g
Frequency   Cell-cycle period, t_{cc} (days)

that generated it). We therefore developed a robust quantitative approach where cell-cycle attributes estimated from H2B-GFP experiments were embodied in (non-Markovian) model simulations, and a subsequent maximum likelihood estimation (MLE) method was applied across the available data sets for each body site to challenge whether each of them was consistent with the SP paradigm (Fig. 6; Supplementary Methods).

Clonal dynamics in esophageal epithelium. In order to explore in vivo clonal dynamics, we began by studying a lineage tracing data set from mouse esophageal epithelium (Fig. 7a). In this experiment we used a strain (*Lrig1-cre*) in which a tamoxifen-regulated form of *Cre* recombinase and enhanced green fluorescent protein (EGFP) are targeted to one allele of the *Lrig1* locus[8,23–25]. We found LRIG1 protein was ubiquitously expressed

Fig. 3 Analysis of cell proliferation in epidermis and esophageal epithelium. a Protocol: $R26^{M2rtTA}/TetO\text{-}H2BGFP$ mice were treated with doxycycline (Dox) to induce H2BGFP expression (green). Following Dox withdrawal, H2BGFP transcription ceases and protein levels dilute with cell division. **b** H2BGFP fluorescence was quantified in non-mitotic basal cell nuclei in optical sections of the basal layer of wholemounts. Scale bar, 20 µm. **c** Representative confocal z stacks of the esophageal basal layer showing H2BGFP (green) and pan-leukocyte marker CD45 (red). Images are representative of a total of 15 fields of view from 3 individual biologically independent mice at 0, 7, and 12 days and 10 fields of view from two individual mice at 18 days. Infrequent label-retaining cells (LRCs) (arrowhead) are positive for CD45 (insert; blue: DAPI). Scale bars, 20 µm. Histograms show keratinocyte H2BGFP intensity for each time point (in green; bottom panels) with mean ± s.e.m. values from each field of view. Best fits for the SP model with exponential- (gray) or gamma-distributed cell-cycle periods (orange lines) are shown. Raw values for H2B-GFP intensity are given in Supplementary Data 2. **d** Outcome of Silverman's unimodality test applied to individual-cell H2BGFP distributions at 18 days in esophagus and epidermis (analyses are separated per animal; this test is effectively two tailed, no multiple-testing corrections are made on p values, exact p values are given in Supplementary Data 3)[46]. A single tissue in a single animal was found to be bimodal, in this case due to variability between fields of view where cells differed by a single round of division (see Fig. 5c). **e** SP-model simulations using different distributions of cell-cycle times t_{cc} (EXP: exponential; GAM: Gamma) with the same average division rate (blue vertical line), the Gamma-shaped distributions predict a more homogeneous dilution. **f** Time course in the H2BGFP intensity distributions from esophageal epithelium (normalized to average keratinocyte intensity at time 0). Green boxplots: experimental data from $n = 3$ biologically independent mice at 0, 7, and 12 days and 2 at 18 days. Centre line of box is median value, box indicates 25th and 75th centiles and whiskers indicate minimum and maximum values. The computed average division rate $\langle\lambda\rangle = 2.9/\text{week}$; solid black line. Gray region: range of H2BGFP intensities predicted from models assuming exponentially distributed cell-cycle times (interquartile range, inner dashed black lines). Light orange region: range of H2GFP intensities inferred with a gamma cell-cycle time distribution (interquartile range in dark orange, delimited by inner dashed orange lines). **g** Most-likely (gamma) shapes for the distribution of the cell-cycle period of esophageal keratinocytes, estimated from fits to the H2BGFP-dilution data. A conservative solution (in dark orange) is used for further inference. Vertical blue line: mean cell-cycle period.

in the basal layer of esophageal epithelium in wild type mice (Supplementary Fig. 3A). Consistent with this finding, in *Lrig1-cre* animals, EGFP, reporting *Lrig1* transcription was detected in $94 \pm 0.3\%$ (s.e.m.) of basal cells (Supplementary Fig. 3B). These observations indicate *Lrig1* is widely expressed in the proliferative compartment of esophageal epithelium and is suitable for lineage tracing of proliferating esophageal keratinocytes (Supplementary Methods).

To track the fate of basal cells, *Lrig1-cre* mice were crossed with the *Rosa26^flConfetti/wt* (*Confetti*) reporter strain which labels cells with one of four possible fluorescent proteins (green, GFP, cyan, CFP, yellow, YFP, or red RFP) after recombination (Supplementary Fig. 3c, d)[26,27]. In some *Cre* inducible mouse lines, reporter expressing clones can appear without induction with Tamoxifen. However, no fluorescent protein expression was found in adult uninduced *Lrig1-cre/Confetti* mice (Supplementary Data 5)[28]. Next, cohorts of *Lrig1-cre/Confetti* animals were treated with a low dose of Tamoxifen that resulted in labeling of only 1 in 300 ± 106 (mean ± s.e.m.) basal cells at 10 days post induction. Clones containing one or more basal cells were imaged in esophageal epithelial wholemounts from at least three mice at multiple time points over 6 months following induction (Fig. 7b; Supplementary Fig. 4a; Supplementary Data 5). Only CFP, YFP, and RFP expressing clones were counted because of *Lrig1*-driven GFP expression in all basal cells.

The pooled *Confetti* clone data set displayed several important features, which were recapitulated by clones labeled with each individual reporter. No statistically significant differences were observed between CFP, YFP, and RFP clone-size distributions at each time point (see Methods). The density (clones/area) of labeled clones decreased progressively, consistent with clone loss through differentiation, while the number of basal and suprabasal cells in the remaining clones rose (Fig. 7c; Supplementary Fig. 4B, C). The proportion of labeled basal cells remained constant during the experiment, indicating the labeled population was self-maintaining over a 6-month period, consistent with labeled cells being a representative sample of all proliferating cells in the homeostatic tissue (Fig. 7c). At late time points, the clone-size distribution scaled with time. This means that if, for example, time doubles, not only the average clone-size shape and breadth of the clone-size distribution also double. More formally, the probability of seeing clones larger than x times the average clone size became time-invariant, following a simple exponential $f(x) = e^{-x}$ (Supplementary Fig. 4D)[29]. Collectively, these features are

hallmarks of neutral competition, in which clonal dynamics result from stochastic cell fates, with an average cell division generating one proliferating and one differentiating daughter cell, a scenario consistent with the SP model (Supplementary Methods)[3,29,30].

The measurement of the average cell division rate (λ) and inference of the cell-cycle time distribution constrain the fitting of lineage tracing data, providing a stringent test of the candidate SP model (Supplementary Figs. 1A–C, 2E). Within this paradigm unknown parameters are the probability of a progenitor cell division generating two dividing (PP), or two differentiating (DD) daughters (r), and the stratification rate (Γ), which in homeostasis sets the fraction of progenitor cells in the basal layer (ρ) (Fig. 6). Our technique for identifying the most appropriate cell-cycle distribution coupled with an MLE grid search estimated parameter values that gave an excellent fit with the clone-size distributions at both early and late time points for the *Lrig1/confetti* data set (Fig. 7d–e; Table 1). The model predictions were within the 95% confidence interval of the measured proportion of clones of a given size at each time point in 27/28 cases. To quantify the quality of the fit, we calculated both the determination coefficient between the model prediction and measured clone sizes, averaged across all time points, R_T^2, and the standard error of the fit, S_T, a measure of the standard deviation between the model estimates and the experimental data, averaged over all time points. For the fit of the SP model to the *Lrig1/Confetti* data set, $R_T^2 = 0.93$, $S_T = 4.3$. Values of R^2 and S for experimental data at each time point are given in Supplementary Data 4.

Next, we applied the same approach to an independent, published lineage tracing data set from esophageal epithelium where clones were labeled with YFP by *Cre* expressed from an inducible *Cyp1a1* (*Ah*) promoter in *AhYFP* mice[5]. Parameter values very similar to those from the *Lrig1/confetti* experiment gave predictions from the SP model within the 95% CI for all 49 points in the experimental data set (Fig. 7d, e; Supplementary Fig. 4e; Table 1; Supplementary Data 4; Supplementary Methods). Quantifying the quality of fit, we found $R_T^2 = 0.98$, $S_T = 2.8$. We noted that including the cell-cycle time constraints resulted in an improved agreement with early time point clone sizes compared with the original publication ($R_T^2 = 0.97$, $S_T = 3.3$), where cell-cycle time distributions were assumed exponential (see Supplementary Data 4 for detailed goodness-of-fit statistics)[5].

As a further validation, we tested the predictions of the SP model against a third, more limited data set from *Krt15-cre^{PR1}* $R26^{mT/mG}$ mice in which a red-to-green fluorescent reporter was used with
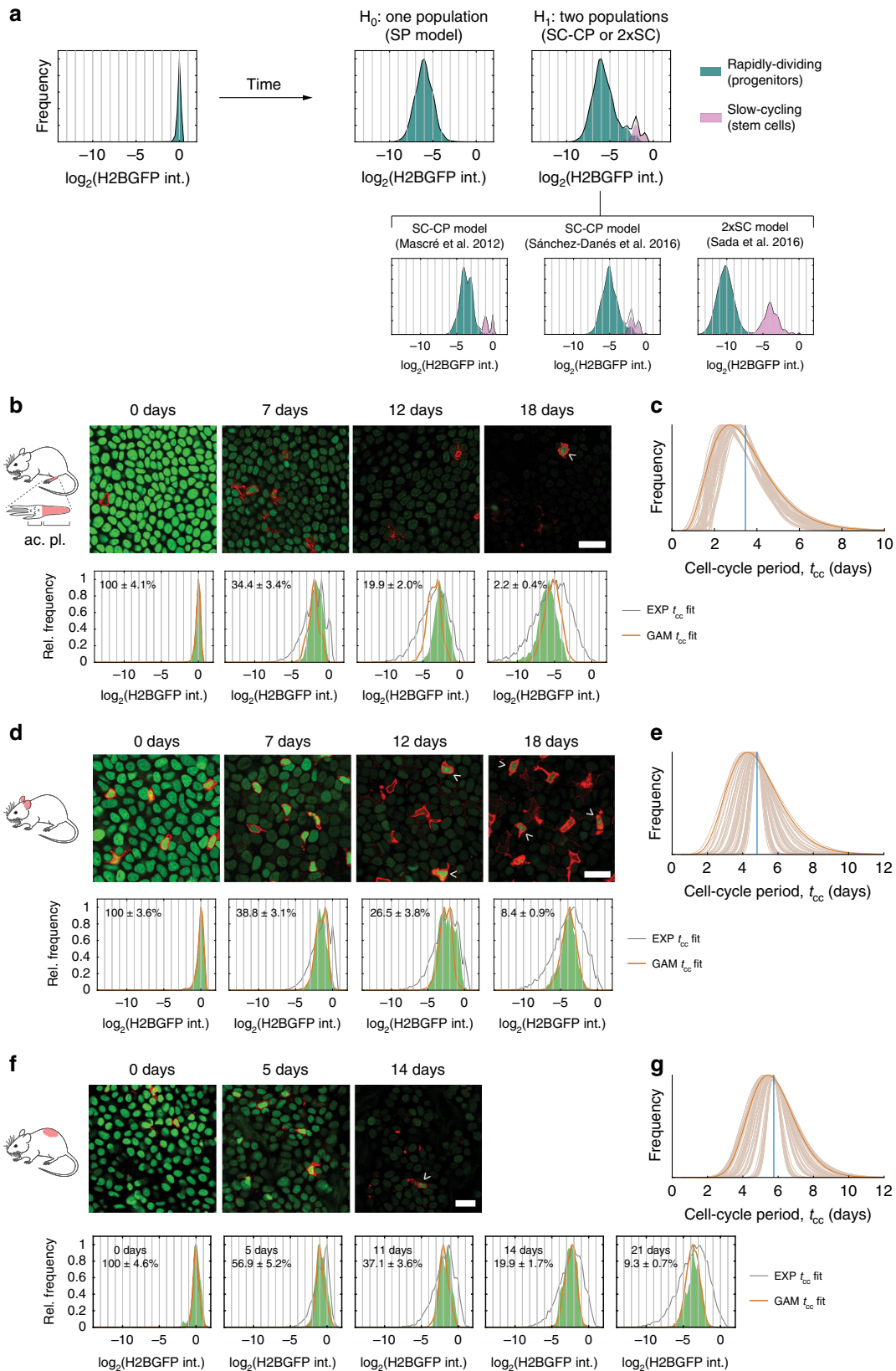
inducible *Cre* expressed from a *Krt15* promoter (Supplementary Methods)[31]. Although the SP paradigm was criticized by these authors, it yielded an adequate fit ($R_T{}^2 = 0.91$, $S_T = 2.6$) with their own data over the experimental time course (Supplementary Fig. 4f; Supplementary Data 4). The consistent agreement of the SP model

to three independent lineage tracing data sets using different combinations of transgenic *Cre* and reporter alleles strongly supports the conclusion that the esophageal epithelium is maintained by a single-progenitor population and argues for the reliability of our parameter estimates.

Fig. 4 H2BGFP-dilution and cell-cycle inference in skin epidermis at different sites. a Theoretical distributions of individual-cell H2BGFP intensities expected after 3 weeks dilution under the SP, SC-CP and 2xSC scenarios assuming gamma-distributed cell-cycle periods. Simulations considered an average division rate for stem cells 4× slower than for progenitors in SC-CP and 2xSC (top panels). Predictions of each model using published parameters are shown below. All theoretical SC-CP or 2xSC scenarios represented in this figure were significant by 6 different unimodality tests (see Supplementary Data 3). **b, d, f** Representative confocal z stacks of hindpaw (plantar), ear and dorsal epidermal basal layer, respectively, from $R26^{M2rtTA}/TetO$-$H2BGFP$ mice, showing H2BGFP (green) and immunostaining for pan-leukocyte marker CD45 (red). Analysis of hindpaw epidermis was confined to the posterior, plantar region (pl.), excluding the acrosyringia (ac.), cartoon. Label-retaining cells (LRCs) are CD45+ leukocytes (arrowheads). Scale bars, 20 μm. Images shown are representative of a total of 18 fields of view from 3 mice at 0 and 12 days, 14 fields of view from 3 mice at 7 and 18 days in (**b**), 18 fields of view from 3 mice at 0, 7 and 12 days and 12 fields of view from 2 mice at 18 days in (**d**), and 20 fields of view from 4 mice at 0 and 11 days and 17 fields of view from 4 mice in (**f**). Lower panels: Individual-keratinocyte H2BGFP intensity levels (in green) with mean ± s.e.m. values from different fields of view. Best fits for the SP model with exponential- (gray) or gamma-distributed cell-cycle periods (orange lines) are shown. See Supplementary Data 2 for raw intensity values and summary statistics. **c, e, g** Best estimates for the (gamma) distribution of the keratinocyte cell-cycle times in hindpaw, ear, and dorsal epidermis, respectively, estimated from fits to H2BGFP-dilution data. Conservative solutions (in dark orange) are used for further inference. Vertical blue lines: average cell-cycle period per site: $\langle\lambda\rangle = 2.0, 1.5, 1.2$/week for paw (**c**), ear (**e**), and dorsum (**g**), respectively.

Clonal dynamics in skin epidermis. We next investigated clonal dynamics in the epidermis through available lineage tracing data sets from the typical interfollicular epidermis of the mouse hindpaw (plantar), ear and back (dorsum). Applying the MLE approach constrained by the cell-cycle time analysis at each body site yielded slightly improved fits of the SP model to data from $Axin2$-$cre^{ERT}R26^{Rainbow}$ animals in paw epidermis ($R_T^2 = 0.98$, $S_T = 2.8$) and also ear epidermis ($R_T^2 = 0.97$, $S_T = 3.5$) and dorsal interfollicular epidermis ($R_T^2 = 0.94$, $S_T = 5.0$) from $AhYFP$ mice compared with those fits reported in the original publications ($R_T^2 = 0.93$, $S_T = 5.16$; $R_T^2 = 0.97$, $S_T = 3.52$; $R_T^2 = 0.92$, $S_T = 6.01$; respectively) (Fig. 8; Supplementary Methods, Supplementary Data 4)[11,22,32]. Despite differences in average keratinocyte division rates across territories ($\lambda \approx 2.0, 1.5, 1.2$/week for plantar hindpaw, ear and dorsum, respectively), all analyzed regions share comparable intermediate proportions of progenitor basal cells ρ (~55%, the rest corresponding to differentiating basal cells) and a predominance of asymmetric cell divisions (i.e., low inferred values for the probability of symmetric division, $r < 0.25$) (Table 1; Supplementary Data 4).

Particularly relevant are the implications for the mode of keratinocyte renewal in back skin, as a previous work claims that two stem cell populations dividing at different rates coexist at this site (2xSC model)[20]. This argument was supported by a quantitative analysis of H2BGFP-dilution patterns in $Krt5^{tTA}/pTRE$-$H2BGFP$ mice, a system that differs from that we use above in that mice are treated with Dox to suppress H2B-GFP expression instead of using it as activator (Supplementary Fig. 5A)[20]. However, in that publication, which rejected the SP model, exponential distributions for cell division/cell stratification rates were assumed. Here we have shown this to be inappropriate for the short time scale of the experiment (Supplementary Methods). Our computational reanalysis, constrained by cell-cycle time distributions demonstrated the SP model gave as good a fit to the $Krt5^{tTA}/pTRE$-$H2BGFP$-dilution data as the more complex 2xSC hypothesis (SP $R_T^2 = 0.85$, $S_T = 0.06$ vs. 2xSC $R_T^2 = 0.87$, $S_T = 0.05$ for basal layer; SP $R_T^2 = 0.78$, $S_T = 0.07$ vs. 2xSC $R_T^2 = 0.79$, $S_T = 0.06$ for spinous layer) (Supplementary Fig. 5B; Supplementary Methods; Supplementary Data 4). Indeed, the inferred parameter values from the $AhYFP$ mouse back skin epidermis proved robust, providing good fits to another lineage tracing data set from the same body site in $Lgr6$-$eGFPcre^{ERT}Rosa26^{flConfetti}$ mice ($R_T^2 = 0.96$, $S_T = 2.39$) (Supplementary Fig. 5C; Supplementary Methods; Supplementary Data 4)[33].

Finally, we turned to revisit clonal dynamics in the mouse tail epidermis. Previous studies of tail have argued that the hierarchical SC-CP paradigm applies to proliferating cells in the interscale areas while the SP paradigm describes behavior in the scale regions (Fig. 5a)[4,21]. These claims were primarily supported by the observation of LRCs in the interscale region in H2BGFP-dilution experiments in $Krt5^{tTA}/pTRE$-$H2BGFP$ mice. However, our quantitative reanalysis of this data set showed the SP-model fits the reported H2BGFP intensity histograms over time as well as the SC-CP model (SP $R_T^2 = 0.89$, $S_T = 0.04$ vs. SC-CP $R_T^2 = 0.89$, $S_T = 0.04$) (Supplementary Fig. 6A; Supplementary Data 4)[21]. Even though we cannot discard the possibility of a subpopulation of slow-cycling stem cells in the tail, such cells would seem to be rare in interscale epidermis (Supplementary Data 1). We noted that a large proportion of the rare LRCs were identified as CD45 expressing leukocytes in our data set (Fig. 5b; Supplementary Data 1). Further analysis argued that there was no conflict between the reported tail lineage tracing data and the SP model (Supplementary Fig. 6B-H; Supplementary Methods; Supplementary Data 4).

Discussion

Overall, we find that combining cell-cycle distribution analysis with lineage tracing argues mouse esophageal epithelium and epidermis are maintained by a single population of progenitor cells, with the sole possible exception of the interscale compartment of tail skin. The quality of the fit of the SP model to the data is equivalent to or exceeds that of more complex models, rendering the need to invoke additional cell populations redundant. The nine lineage tracing data sets analyzed include a variety of Cre and reporter strain combinations, and are all consistent with the SP model. In addition, live imaging studies of the epidermis are consistent with a single proliferating cell population maintaining the tissue[34].

Quantitative analysis of cell proliferation in the different tissue types identifies further constraints that must be considered by researchers exploring the appropriateness of alternative models. The original SC-CP and 2xSC models invoked 12% and 30% of basal-layer keratinocytes constitute slow-cycling stem cells, respectively[4,20]. Histone dilution experiments have allowed us to make strong statements about the nature of any proposed second population. For each body site, with the exception of tail interscale epidermis, no keratinocyte label-retaining cells were detected in over 2000 cells imaged at each location. It follows that any slow-cycling stem-cell population must have substantially fewer than one slow-cycling cell per thousand basal keratinocytes to be compatible with observations reported here, making it unlikely that such slow-cycling cells will make a detectable contribution to tissue homeostasis. The hypothesis that two subpopulations exist, but that they both divide at a similar rate, is hard to sustain in the face of the close agreement of the simpler SP model across all the analyzed data sets.
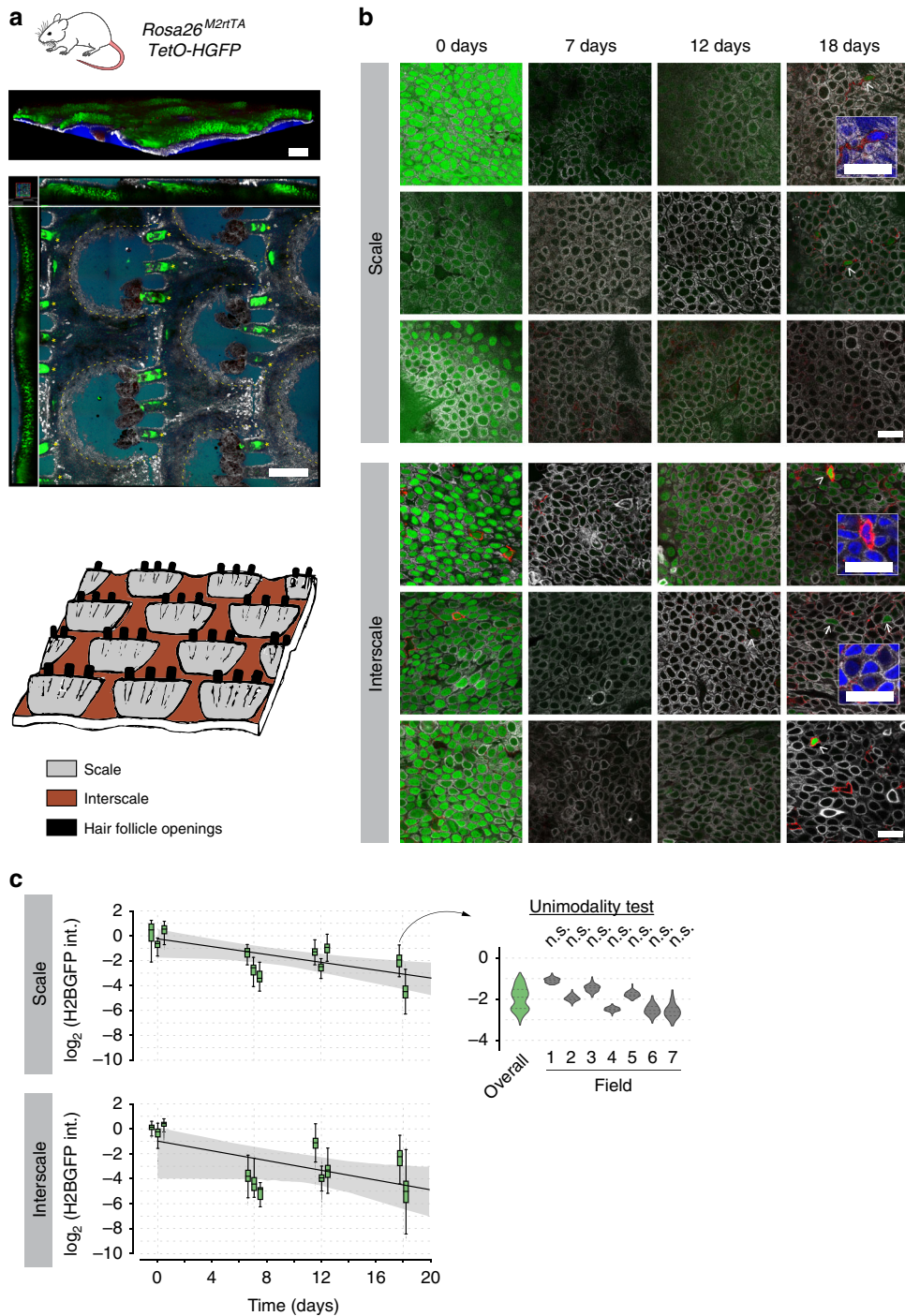
Fig. 5 H2BGFP-dilution analysis in tail epidermis. a Structure of the mouse tail epidermis. 3D reconstruction of confocal *z* stacks (top panel) and orthogonal xyz view (mid panel). Apical blister-shaped regions (scale) alternate with deeper regions (interscale) (boundaries delineated by dotted lines), being arranged in lines separated by hair follicles (asterisks). Green: H2BGFP expression; white: KRT14 immunostaining (as a marker of basal layer); red: CD45 immunostaining; blue: DAPI. Scale bars, 200 μm. Bottom panel: Cartoon illustrating the tail skin structure. **b** Representative confocal z stacks of scale and interscale regions of tail epidermis during H2BGFP chase experiments in *R26^{M2rtTA}/TetO-H2BGFP* mice, showing H2BGFP (green), immunostaining for KRT14 (white) and immunostaining for pan-leukocyte marker CD45 (red). Images from the same time point correspond to different mice to illustrate inter-animal variation. Label-retaining cells (LRCs) are highlighted with arrowheads (CD45+ cells) or full arrows (CD45− cells) (details in inserts; blue: DAPI). Scale bars, 20 μm. **c** Time course of basal-layer keratinocyte H2BGFP intensity distributions from scale and interscale regions of tail. Experimental data shown as boxplots per individual biologically independent mouse (intensities normalized to average keratinocyte intensity at time 0, raw H2B-GFP intensity values are given in Supplementary Data 2). *n* = 3 animals at each time point except 18 days where *n* = 2 mice. Centre line of box is median value, box indicates 25th and 75th centiles and whiskers indicate minimum and maximum values. Solid black lines: average H2BGFP-dilution rates (within 95% CI limits—shaded gray areas—where the value of *λ* = 1.2/week reported by Mascre et al.[4] falls). Insert: Detail of H2BGFP intensity distributions separated per field of view for the single tissue found to be bimodal in the overall, per-animal modality test (see Fig. 3d). Analyses per field of view all resulted non-significant (unimodality).

Table 1 Parameter values inferred for progenitor cell behavior in different murine epithelial regions as derived from quantitative lineage tracing.

Tissue	Experimental model	Reference	min tcc (days)	Division rate, λ (/week)	Symmetric division prob., r	% of progenitor cells, ρ	Stratification rate, Γ (/week)
Esophagus	*Lrig1-eGFP-Cre^ERT/R26-^flConfetti*		0.5	2.9 (2.7; 3.0)	0.10 (0.07; 0.15)	65 (50; 96)	5.4 (2.9; 69.6)
	Ah-Cre^ERT/R26^flEYFP	5	0.5	2.9 (2.7; 3.0)	0.06 (0.04; 0.10)	56 (50; 89)	3.7 (2.9; 23.5)
Paw Epidermis	*Axin2-Cre^ERT2/R26-Rainbow*	11	1	2.0 (1.7; 2.3)	0.14 (0.12; 0.17)	53 (49; 58)	2.3 (1.9; 2.8)
Ear Epidermis	*Ah-Cre^ERT/R26^flEYFP*	32	1	1.5 (1.2; 1.7)	0.04 (0.03; 0.06)	54 (47; 72)	1.8 (1.3; 3.9)
Back Epidermis	*Ah-Cre^ERT/R26^flEYFP*	22	2	1.2 (1.1; 1.3)	0.04 (0.03; 0.07)	61 (55; 76)	1.9 (1.5; 3.8)

Parameter values indicated correspond with the maximum likelihood estimate (MLE), values in parentheses are 95% confidence bounds (see Supplementary Methods for details).
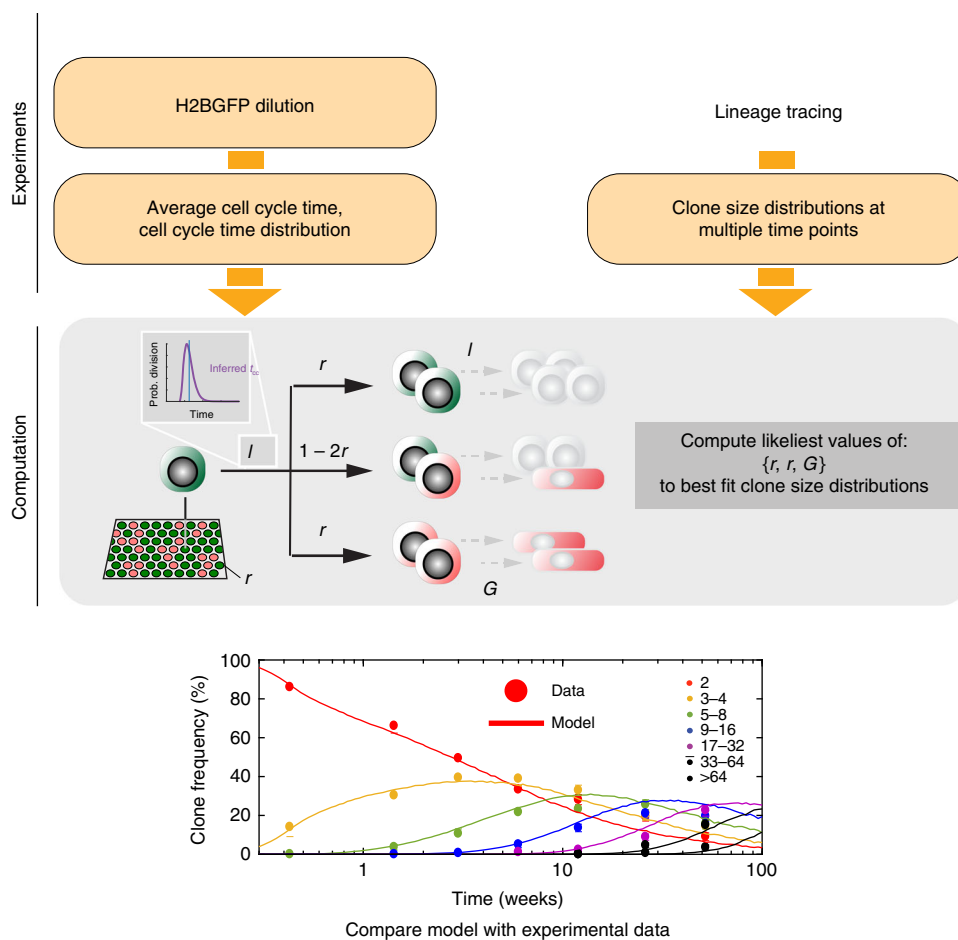
Fig. 6 Method for single-progenitor model testing and parameter inference. Method to single-progenitor model testing and infer model parameters. Orange boxes indicate experiments and resulting data, gray box computational model and parameter estimation. Italics indicate parameters in the SP model. The multimodality testing of H2B-GFP data showed that there is a single population dividing at the same average rate in epidermis and esophagus, consistent with the SP model (Fig. 3d). To test the SP model, the average cell-cycle time (λ) and cell-cycle time distribution were inferred from H2B-GFP experiments. These values are used in computational analysis to estimate the values of the other parameters in the SP model, the proportion of progenitor cells in the basal layer ρ, the proportion of symmetric cell division outcomes r, and the stratification rate of differentiating cells leaving the basal cell layer (Γ). Multiple sets of values for the unknown parameters were tested. For each set of unknown parameter values 100,000 progenitor-derived clones were simulated (lines) and inferred clone-size distributions compared with experimental ones (points) obtained from lineage tracing. The likeliest sets of parameter values were obtained by maximum likelihood estimation for each linage tracing data set. The quality of the fit was assessed by determining whether the simulated values lie within the 95% confidence interval of the experimental clone-size measurements at each time point.

The improved resolution of parameter estimates identifies differences in cell division rates across the epidermis and the esophagus (Table 1). Proliferating cells divide rapidly in the esophageal epithelium, on average every ~2.4 days (similar to keratinocyte turnover rate in oral mucosa), while progenitor cells cycle comparatively slower in the epidermis, on average between 3.5 and 6 days depending on body site[35]. However, our study suggests individual cell-cycle periods are tightly controlled, showing little variation around average division rate, per territory.
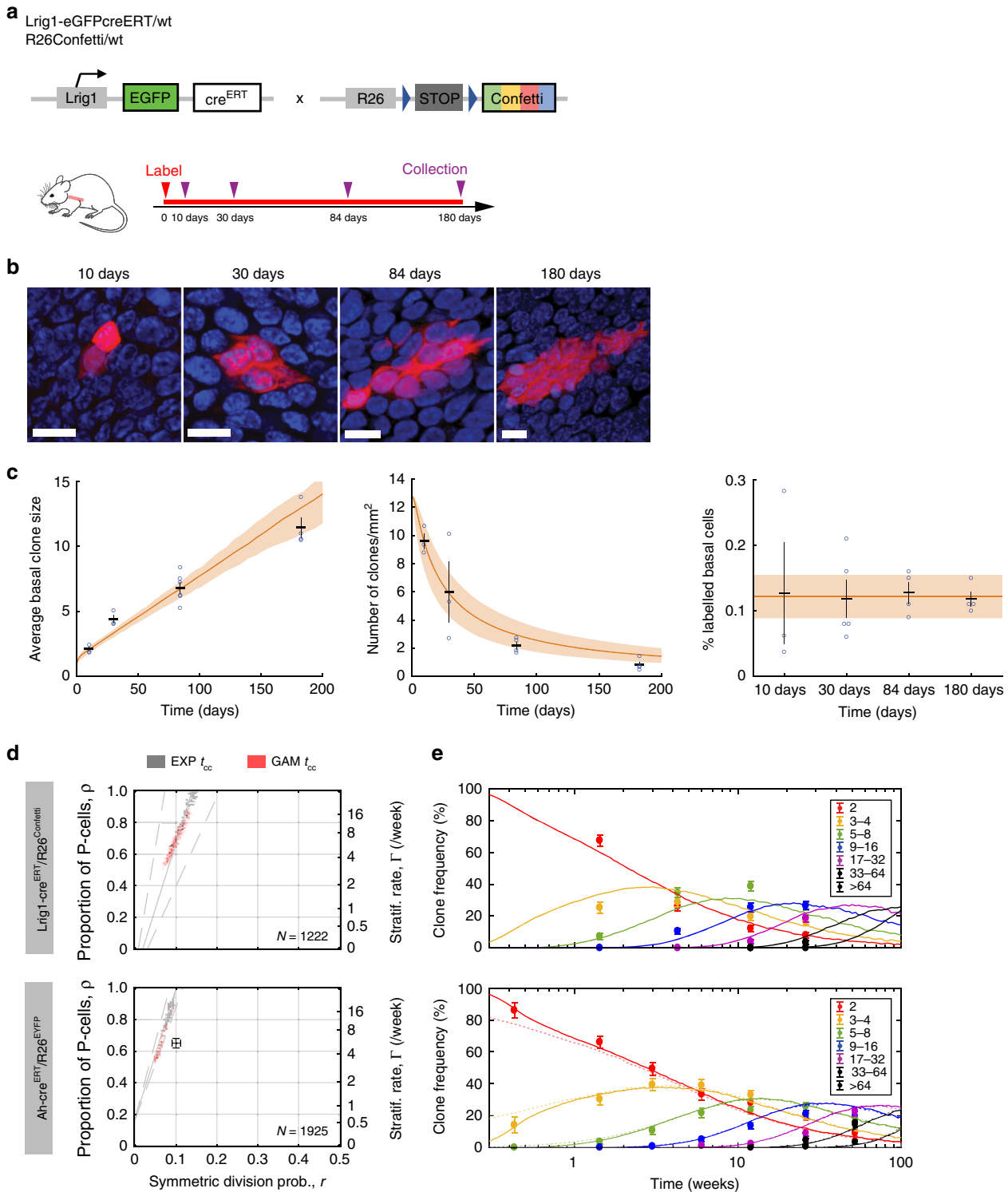
The proportion of progenitor cells in the basal layer and the probability of symmetric cell division outcomes (r) are similar across body sites (Table 1). The insight that a substantial proportion of cells in the basal layer will proceed to differentiate rather than divide will be important for the interpretation for the growing body of single-cell RNA sequencing data in these tissues[33,35]. In addition, the low values of r we identify give insight into the basis of cell fate determination (Fig. 9). In principle, if every basal cell divides or differentiates with equal probability, as proposed by Leblond, r will be 0.25, as expected from any pair of uncorrelated basal cells[1]. However, this scenario is excluded by

our analysis. Instead, the consistent values of $r < 0.25$ indicate the fate of sister cells is preferentially anti-correlated. This phenomenon can be associated to local coordination of neighboring cell stratification and division events[36]. Our results argue that anti-correlation of sister cell fates applies generally in the epidermis and esophagus, pointing to common mechanisms of keratinocyte cell fate regulation.

The single-progenitor model captures the average behavior of progenitor cells during homeostasis. However, epithelia are frequently subject to wounding. To repair the tissue requires a temporary imbalance in cell fate, with the progenitors close to the

Fig. 7 Quantitative lineage tracing in esophageal epithelium. a Protocol: clonal labeling was induced in $Lrig1\text{-}eGFPcre^{ERT/wt}$ $R26^{flConfetti/wt}$ mice and samples analyzed at different times from 10 to 180 days post induction, as single labeled cells develop into clones. See Supplementary Data 5 for source data for panels (**c**) and (**e**). **b** Rendered confocal z stacks of the esophageal basal layer showing typical RFP clones (red) at the times indicated. Blue is DAPI. Scale bars, 10 μm. Images are representative of 104 RFP clones (10 days), 75 RFP clones (30 days), 106 RFP clones (84 days), and 274 RFP clones (180 days). **c** Quantitative characteristics of the labeled clone population over time: average basal-layer clone size (i.e., mean number of basal cells/surviving clone) (left panel), average density of labeled clones in the basal layer (middle panel), average fraction of labeled basal cells at the indicated time points (right panel). Observed values are shown in individual biologically independent mice (blue circles, n mice = 3 at 10 and 30 days, 6 at 84 days, and 4 at 180 days) with error bars (black) indicating mean ± s.e.m. of all mice at each time point. A total of 300 or more clones was quantified at each time point. Orange lines: SP-model fit (shaded area corresponds with 95% plausible intervals). Orange line and shading in last panel show mean and s.e.m. across all time points (from $n = 16$ mice), which is consistent with homeostatic behavior. **d** SP-model parameter inference on $Lrig1$- and $Ah\text{-}Cre^{ERT}$ driven lineage tracing data sets from esophagus[5]. Parameter estimates are affected by the underlying modeling assumptions on the cell-cycle period, whether default exponential cell-cycle time distributions were considered (solutions in gray) or realistic gamma distributions implemented, as inferred from the cell-proliferation analysis (solutions in orange). Regions within the dashed gray lines fall consistent with the predicted ρ/r ratios from the linear scaling of the average clone size. The total number of clones counted in each data set is displayed in the corresponding graph and previous parameter estimates given in ref. [5] shown as black error bars. **e** Experimental $Lrig1$- and $Ah\text{-}Cre^{ERT}$-derived basal-layer clone sizes from ref. [5] (dots with error bars indicating the standard error of a proportion) fit well with the SP model, with gamma-distributed cell-cycle times (lines; prediction from maximum likelihood estimation). Dim dashed lines: fits from ref. [5]. Frequencies for each clone size (basal cell number) are shown in different colors. $n = 3$ biologically independent mice at each time point except for the $Lrig1/confetti$ where $n = 6$ mice at 12 weeks and 4 mice at 26 weeks.

wound producing an excess of progenitor over differentiating daughters on average. This occurs as part of a coordinated set of responses that includes cell migration and altered cell differentiation[5,37,38]. Once the epithelial defect is resolved, the progenitors revert to homeostatic balance. In esophageal epithelium and the plantar epidermis, wound repair is achieved by progenitors alone[5,11]. In the epidermis at other sites, cells migrating from other proliferative compartments, the hair follicles and sweat ducts, may also contribute to wound healing[9,10,39,40]. The ability to transiently increase the likelihood of progenitors generating proliferating progeny provides a rapid and robust response to injury. The down side of this adjustable progenitor fate is that it may be subverted by mutations acquired during tissue aging, leading to mutant clonal expansions that may undergo malignant transformation[22,41–43].

How might these findings in mice relate to homeostasis human epidermis? Human skin differs from that of mice with many more epidermal cell layers and undulates in thickness at most body sites creating folds called rete ridges and dermal papillae[44]. Nevertheless, a population of cells with balanced stochastic cell fate generating equal proportions of proliferating and differentiating cells has been identified in a live imaging study of human keratinocytes in primary culture[37]. In vivo lineage tracing in humans is not feasible. However, human epidermis has been grafted onto immune compromised mice and injected with lentiviral vectors carrying fluorescent protein reporters. When the resulting clones were imaged 6 months later they were found to vary widely in size and shape and arise from any point in the basal layer, both in rete ridges and dermal papillae[45]. These findings are consistent with the single-progenitor paradigm, but cannot provide quantitative challenge to the model available in mice.

The lineage tracing approaches considered above have been enriched by live imaging studies of mouse epidermis[34,36]. Whilst lineage tracing resolves the average behavior of a population of proliferating cells over many cell generations, live imaging allows the fate of individual cells to be resolved. Insights gained from live imaging include showing that cell fate is stochastic, the probability of generating progenitor and differentiated daughters is equal and that the fate of cells is not coordinated across cell generations, all of which are key features of the SP model[34].

We conclude that the single-progenitor model is consistent with a large body of lineage tracing and cell-cycle data collated from multiple studies and identifies the behavior of proliferating cells that underpins epidermal and esophageal epithelial homeostasis.

Methods

Animals. All experiments were conducted according to the UK Home Office Project Licenses 70/7543, P14FED054 or PF4639B40. Male and female adult mice aged 3–18 months were used for in vivo experiments. Animals were housed in individually ventilated cages and fed on standard chow and maintained in SOPF health status.

Doubly transgenic, $Lrig1\text{-}eGFPcre^{ERT/wt}$ $R26^{flConfetti/wt}$ mice on a C57/Bl6N background were generated for lineage tracing studies in esophageal epithelium, by crossing $Lrig1\text{-}eGFP\text{-}ires\text{-}cre^{ERT2}$ mice[8] onto a $Rosa26^{flConfetti}$ multicolor reporter line[26]. Transcription of the Cre recombinase-mutant estrogen receptor fusion protein (Cre^{ERT}) is under the control of an endogenous allele of $Lrig1$. Following induction with tamoxifen, Cre^{ERT} protein internalizes into the nuclei and excises a $LoxP$-flanked "STOP" cassette resulting in the expression of one of the four Confetti fluorescent reporters (YFP, RFP, CFP, or GFP). $R26^{M2rtTA}/TetO\text{-}H2BGFP$ mice, doubly transgenic for a reverse tetracycline-controlled transactivator (rtTA-M2) targeted to the $Rosa26$ locus and a HIST1H2BJ/EGFP fusion protein (H2BGFP) expressed from a tetracycline promoter element, were used for label-retaining experiments[5,22]. H2BGFP expression is induced by treatment with doxycycline (Dox) and dilution of H2BGFP protein content can be chased upon Dox withdrawal. All animals were induced at 8–12 weeks age. Cohorts of at least two or three animals per time point were culled and esophagus and/or skin epidermis collected for analysis.

Wholemount preparation and immunostaining. Esophageal epithelium wholemounts for lineage tracing were prepared as follows: The esophagus was cut longitudinally and the middle two-thirds of the tract was incubated for 3 h in 5 mM EDTA in PBS at 37 °C. The epithelium was then peeled away from the underlying submucosa, stretched and fixed for 30 min in 4% paraformaldehyde in PBS. Samples were stored in PBS at 4 °C until subsequent analysis. Skin pieces of ~0.5 cm² were cut and incubated for 1 h in 5 mM EDTA in PBS at 37 °C. Skin epidermis was then peeled away using fine forceps and processed as described above for the esophageal epithelium.

For staining, wholemount samples were incubated in Permeabilization Buffer (PB) (0.5% BSA, 0.25% Fish Skin Gelatin (FSG), 0.5% Triton X-100/PBS) for 15 min at room temperature (RT), then blocked in 10% goat or donkey serum/PB (according to the secondary antibody used) for 1 h at RT and incubated overnight with primary antibody at 4 °C. Primary antibodies used were Lrig1 antibody (R&D Systems, Cat. AF3688), ITGA6 antibody (clone GoH3, Biolegend, Cat. B204094), Alexa Fluor® 647 anti-CD45 (clone 30-F11, Biolegend, Cat. 103124), Keratin 14 antibody (clone Poly19053, Biolegend, Cat. 905301). Samples were subsequently washed four times for 30 min in 0.2% Tween-20/PBS and incubated with an appropriate secondary antibody for 3 h at RT. Secondary antibodies used were Goat or Donkey Alexa Fluor 488/546/555/647 (Molecular Probes). A washing step with 0.2% Tween-20/PBS was repeated and samples were incubated for 30 min with DAPI (Sigma-Aldrich) and finally mounted in DAKO Vectashield Mounting Medium with DAPI (Vector Labs).

Dilution of Histone 2B-GFP protein content. $R26^{M2rtTA}/TetO\text{-}H2BGFP$ animals were treated with doxycycline (Dox, 2 mg/ml in drinking water sweetened with 10% blackcurrant & apple) for 4 weeks. Dox was then withdrawn and animals culled at different time points to track H2BGFP florescence dilution. Epithelial wholemounts from esophageal epithelium and skin epidermis were imaged on a Leica TCS SP8 confocal microscope using ×20 or ×40 objectives at 1024 × 1024 resolution, line average 4 and 400 Hz scan speed. Individual-cell H2BGFP
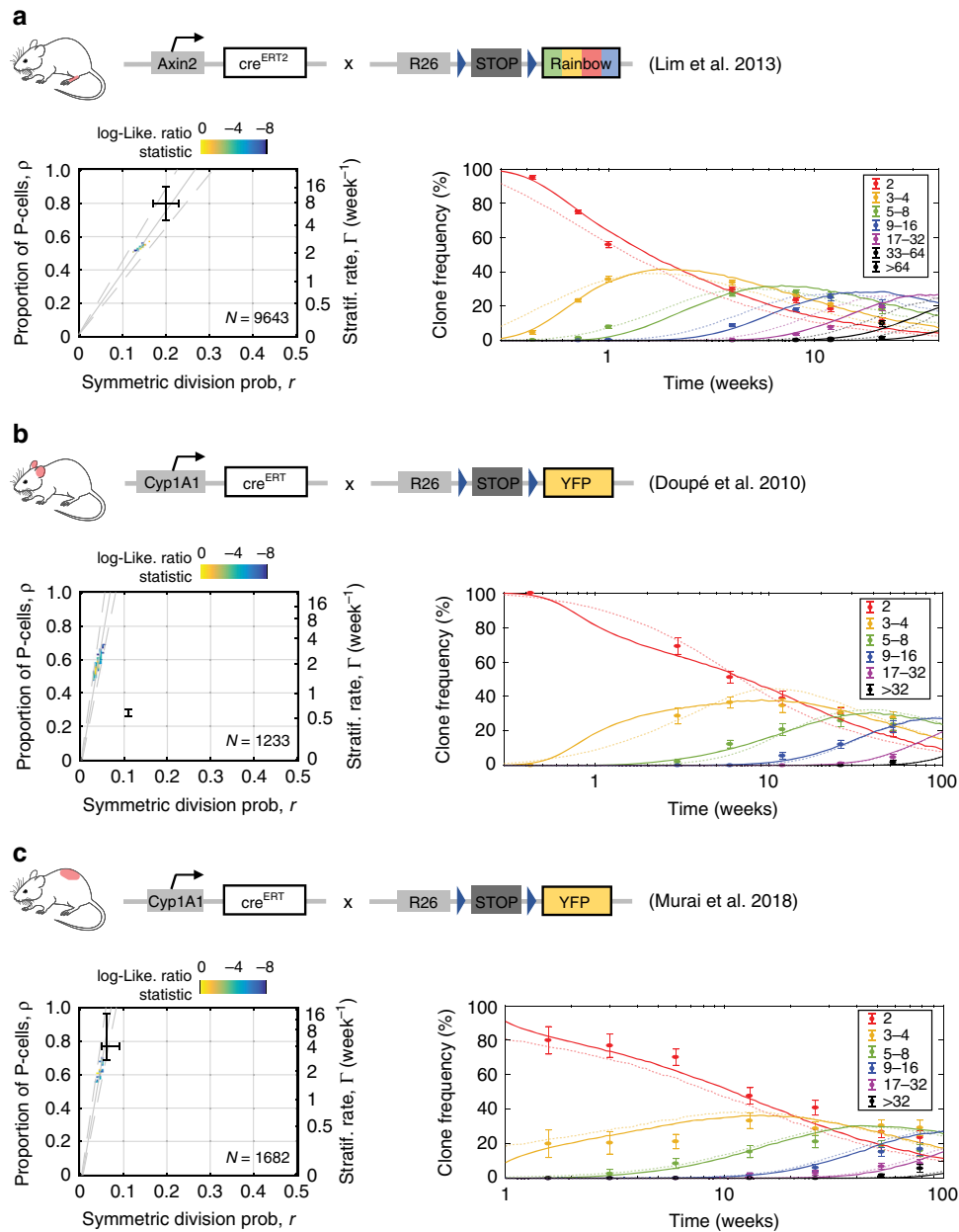
Fig. 8 The single-progenitor model fits clone dynamics in different regions of skin epidermis. a–c Left panels: SP-model parameter inference on lineage tracing data sets from paw epidermis[11] using *Axin2-cre^ERT^R26^Rainbow^* animals (**a**), and ear[32] and dorsal[22] interfollicular epidermis in *AhYFP* mice (**b** and **c**, respectively). Parameter estimates are obtained by MLE based on SP-model simulations constrained by the cell-cycle period distribution inferred from each corresponding cell-proliferation analysis. Regions within the dashed gray lines fall consistent with the predicted ρ/r ratios from the linear scaling of the average clone size. The total number of clones counted in each data set is displayed in the corresponding graph. Black bars are parameter estimates given in the original publications shown, centre is the mid range and bars indicate the maximum and minimum plausible parameter values in simulations in each paper (**a**, **b**, **c**). Right panels: Experimental *Axin2-* (**a**; from ref. [11]) and *Ah-* (**b** and **c**; from refs. [32] and [22]) derived basal-layer clone sizes (dots indicate mean ± standard error of proportion) give an excellent fit with the SP model with gamma-distributed cell-cycle times (lines; prediction from MLE). Dim dashed lines: fits obtained with parameter estimates given in the original publications. Frequencies for each clone size (basal cell number) are shown in different colors. See Supplementary Data 4 for goodness-of-fit statistics.

intensities were determined by image segmentation/nuclear identification, using a semi-automated object-recognition macro (based on the DAPI channel) built in ImageJ, and the process completed by manual curation. Per-cell intensity values given are averaged over all nuclear pixels. All H2BGFP samples were stained for CD45 and positive cells excluded from the analysis.

Lineage tracing. Low-frequency expression of the Confetti reporters in the *Lrig1-eGFP-ires-cre^ERT2^ R26^flConfetti/wt^* mouse esophagus was achieved by inducing 10-week-old animals with intraperitoneal injection of a single dose of 1 mg tamoxifen (100 µl of 10 mg/ml) on two consecutive days[8]. This resulted in a labeling efficiency

of 1 in 301 ± 106 (mean ± s.e.m.) basal cells by 10 days post induction (allowing individual clone tracking without merging). Between three and six mice were culled per time point. Confocal images of immunostained wholemounts were acquired on a Leica TCS SP8 confocal microscope (×10, ×20, and ×40 objectives; typical settings for z-stacks acquisition: optimal pinhole, line average 4, bi-directional scan with 400–600 Hz speed, resolution of 1024 × 1024 pixels).

The number of nucleated basal and suprabasal cells per labeled clone was counted under live acquisition mode. GFP-labeled clones were not scored due to the difficulty of distinguishing them from the constitutive basal GFP expression driven by the *Lrig1* cassette. CFP, RFP, and YFP clones were pooled together for further analysis (histograms (distributions) of basal-layer clone sizes and average
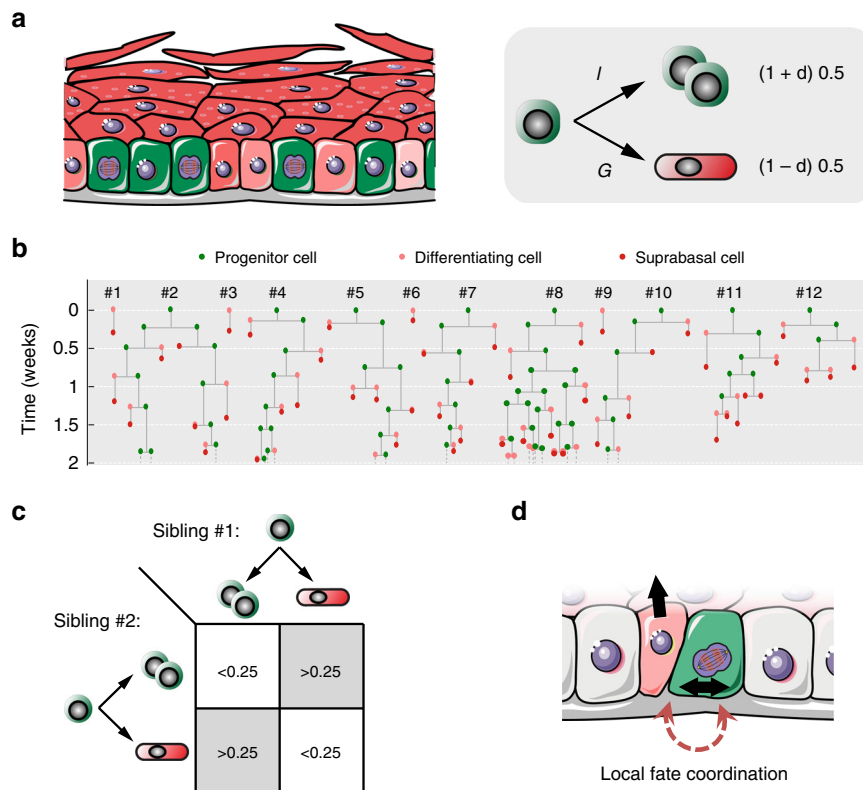
Fig. 9 Cell fate coordination underpins single-progenitor dynamics in epidermis and esophagus. a Epidermis and esophageal epithelium are maintained by a single population of progenitor cells. Left panel: Progenitor cells (in green) share the basal layer with post-mitotic keratinocytes in early stages of differentiation (pale red), which are transiently retained in the basal compartment before stratification. Right panel: Simplified representation of the single-progenitor model focusing on individual basal cell fates. Epithelial cell dynamics are dominated by stochastic but skewed fates through spatial coordination between neighboring or sibling cells. Individual basal cells undertake dichotomic decisions: they have the potential to divide but can alternatively differentiate exiting the basal layer. Both probabilities are balanced (50–50%) across the entire proliferating cell population, but can be decompensated or skewed for individual cells depending on their local niche, as reflected with the parameter δ (which works as a context-dependent modulatory factor). **b** Stochastic progenitor fates explain a scenario of neutral clonal competition dynamics where clones develop into heterogeneous sizes, constrained by cell-cycle time control and fate coordination effects. Displayed is a representative set of epithelial clone dynamics simulated using the parameters inferred for murine esophageal epithelium homeostasis. **c, d** Our results demonstrate that the outcome of sibling keratinocyte cells is commonly biased toward an excess of asymmetric fates where one decides to divide while the other differentiates, in agreement with a single-progenitor model with low values of r ($r < 0.25$).

number of basal cells). A total of 300, 315, 302, and 305 labeled clones from 3, 3, 6, and 4 mice at 10, 30, 84, and 180 days post induction, respectively, were quantified. Regarding the time courses in the number of clones per unit area and the proportion of labeled basal cells, only RFP clones were considered given the low, variable induction of the other florescent reporters and their overall small contribution (including these numbers did not alter the conclusions).

Lineage-tracing data from *Ah-cre*[ERT] *R26*[flEYFP] derived clones in esophagus[5], ear[32], and dorsal epidermis[22] were obtained from experimental colleagues (data available upon request). Data on induced *Axin2-cre*[ERT] *R26*[Rainbow] clones in hindpaw[11] and *Lgr6-eGFPcre*[ERT] *R26*[flConfetti] in back epidermis[33] were kindly provided by the authors. Data from lineage tracing in scale and interscale tail epidermis[21] were accessed through the online publication material, while authors were unable to provide original data from ref. [4]. Data on *Krt15-cre*[PR1] *R26*[mT/mG] mouse esophagus[31] were retrieved by digitalizing Fig. 2e and Figure S3B from the original publication. A similar procedure was used to extract *Krt5*[tTA]/*pTRE*-H2BGFP-dilution data from back skin (Fig. 3 from ref. [20]) and tail epidermis (Fig. 3k from ref. [4]).

Mathematical modeling and statistical inference. Model dynamics were simulated using Markovian (Gillespie algorithm) and non-Markovian exact stochastic Monte Carlo methods implemented in Matlab. A maximum likelihood estimation (MLE) approach was followed for parameter inference, except when stated otherwise, and best-fit parameters obtained with 95% confidence intervals based on the likelihood-ratio test (alpha = 0.05). The coefficient of determination (R^2) and the standard error of the fit (S) were calculated for the evaluation of goodness of fit (GoF). In those cases involving data at various time points, GoF values averaged across the different time points (R_T^2 and S_T) are displayed. A comprehensive list of detailed R^2 and S values can be found in Supplementary Data 4. For details of theoretical modeling and computational methods used to infer cell behavior and clonal dynamics, see Supplementary Methods section below.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The authors declare that the experimental data supporting the findings of this study are available within the paper and its supplementary information files.

Code availability
Code used in computational modeling is available in Github: https://github.com/gp10/Piedrafita_etal_SI_code/

References

1. Marques-Pereira, J. P. & Leblond, C. P. Mitosis and differentiation in the stratified squamous epithelium of the rat esophagus. *Am. J. Anat.* **117**, 73–87 (1965).
2. Allen, T. D. & Potten, C. S. Fine-structural identification and organization of the epidermal proliferative unit. *J. Cell Sci.* **15**, 291–319 (1974).
3. Clayton, E. et al. A single type of progenitor cell maintains normal epidermis. *Nature* **446**, 185–189 (2007).
4. Mascre, G. et al. Distinct contribution of stem and progenitor cells to epidermal maintenance. *Nature* **489**, 257–262 (2012).

5. Doupe, D. P. et al. A single progenitor population switches behavior to maintain and repair esophageal epithelium. *Science* **337**, 1091–1093 (2012).
6. Martincorena, I. et al. Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911–917 (2018).
7. Martincorena, I. et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
8. Page, M. E., Lombard, P., Ng, F., Gottgens, B. & Jensen, K. B. The epidermis comprises autonomous compartments maintained by distinct stem cell populations. *Cell Stem Cell* **13**, 471–482 (2013).
9. Ito, M. et al. Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nat. Med.* **11**, 1351–1354 (2005).
10. Lu, C. P. et al. Identification of stem cell populations in sweat glands and ducts reveals roles in homeostasis and wound repair. *Cell* **150**, 136–150 (2012).
11. Lim, X. et al. Interfollicular epidermal stem cells self-renew via autocrine Wnt signaling. *Science* **342**, 1226–1230 (2013).
12. Kamberov, Y. G. et al. A genetic basis of variation in eccrine sweat gland and hair follicle density. *Proc. Natl Acad. Sci. USA* **112**, 9932–9937 (2015).
13. Mackenzie, J. C. Ordered structure of the stratum corneum of mammalian skin. *Nature* **222**, 881–882 (1969).
14. Spearman, R. I. & Hardy, J. A. Ultrastructure of the contrasting types of keratinization seen in the tail epidermis of the laboratory mouse Mus musculus. *Arch. Dermatol Res.* **258**, 33–40 (1977).
15. Gomez, C. et al. The interfollicular epidermis of adult mouse tail comprises two distinct cell lineages that are differentially regulated by Wnt, Edaradd, and Lrig1. *Stem Cell Rep.* **1**, 19–27 (2013).
16. Braun, K. M. et al. Manipulation of stem cell proliferation and lineage commitment: visualisation of label-retaining cells in wholemounts of mouse epidermis. *Development* **130**, 5241–5255 (2003).
17. Alcolea, M. P. & Jones, P. H. Lineage analysis of epidermal stem cells. *Cold Spring Harb. Perspect. Med* **4**, a015206 (2014).
18. Alcolea, M. P. & Jones, P. H. Tracking cells in their native habitat: lineage tracing in epithelial neoplasia. *Nat. Rev. Cancer* **13**, 161–171 (2013).
19. Tumbar, T. et al. Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363 (2004).
20. Sada, A. et al. Defining the cellular lineage hierarchy in the interfollicular epidermis of adult skin. *Nat. Cell Biol.* **18**, 619–631 (2016).
21. Sanchez-Danes, A. et al. Defining the clonal dynamics leading to mouse skin tumour initiation. *Nature* **536**, 298–303 (2016).
22. Murai, K. et al. Epidermal tissue adapts to restrain progenitors carrying clonal p53 mutations. *Cell Stem Cell* **23**, 687–699 (2018).
23. Nakamura, T. et al. LRIG1 inhibits STAT3-dependent inflammation to maintain corneal homeostasis. *J. Clin. Invest.* **124**, 385–397 (2014).
24. Lu, L. et al. LRIG1 regulates cadherin-dependent contact inhibition directing epithelial homeostasis and pre-invasive squamous cell carcinoma development. *J. Pathol.* **229**, 608–620 (2013).
25. Wong, V. W. et al. Lrig1 controls intestinal stem-cell homeostasis by negative regulation of ErbB signalling. *Nat. Cell Biol.* **14**, 401–408 (2012).
26. Snippert, H. J. et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134–144 (2010).
27. Frede, J., Greulich, P., Nagy, T., Simons, B. D. & Jones, P. H. A single dividing cell population with imbalanced fate drives oesophageal tumour growth. *Nat. Cell Biol.* **18**, 967–978 (2016).
28. Jensen, K. B. et al. Lrig1 expression defines a distinct multipotent stem cell population in mammalian epidermis. *Cell Stem Cell* **4**, 427–439 (2009).
29. Klein, A. M. & Simons, B. D. Universal patterns of stem cell fate in cycling adult tissues. *Development* **138**, 3103–3111 (2011).
30. Klein, A. M., Doupe, D. P., Jones, P. H. & Simons, B. D. Kinetics of cell division in epidermal maintenance. *Phys. Rev. E Stat. Nonlin Soft Matter Phys.* **76**, 021910 (2007).
31. Giroux, V. et al. Long-lived keratin 15+ esophageal progenitor cells contribute to homeostasis and regeneration. *J. Clin. Invest.* **127**, 2378–2391 (2017).
32. Doupe, D. P., Klein, A. M., Simons, B. D. & Jones, P. H. The ordered architecture of murine ear epidermis is maintained by progenitor cells with random fate. *Dev. Cell* **18**, 317–323 (2010).
33. Fullgrabe, A. et al. Dynamics of Lgr6(+) progenitor cells in the hair follicle, sebaceous gland, and interfollicular epidermis. *Stem Cell Rep.* **5**, 843–855 (2015).
34. Rompolas, P. et al. Spatiotemporal coordination of stem cell commitment during epidermal homeostasis. *Science* **352**, 1471–1474 (2016).
35. Jones, K. B. et al. Quantitative clonal analysis and single-cell transcriptomics reveal division kinetics, hierarchy, and fate of oral epithelial progenitor cells. *Cell Stem Cell* **24**, 183–192.e188 (2019).
36. Mesa, K. R. et al. Homeostatic epidermal stem cell self-renewal is driven by local differentiation. *Cell Stem Cell* **23**, 677–686.e674 (2018).
37. Roshan, A. et al. Human keratinocytes have two interconvertible modes of proliferation. *Nat. Cell Biol.* **18**, 145–156 (2016).
38. Park, S. et al. Tissue-scale coordination of cellular behaviour promotes epidermal wound repair in live mice. *Nat. Cell Biol.* **19**, 155–163 (2017).
39. Levy, V., Lindon, C., Harfe, B. D. & Morgan, B. A. Distinct stem cell populations regenerate the follicle and interfollicular epidermis. *Dev. Cell* **9**, 855–861 (2005).
40. Levy, V., Lindon, C., Zheng, Y., Harfe, B. D. & Morgan, B. A. Epidermal stem cells arise from the hair follicle after wounding. *Faseb J.* **21**, 1358–1366 (2007).
41. Fernandez-Antoran, D. et al. Outcompeting p53 mutant cells in the normal esophagus by redox manipulation. *Cell Stem Cell* **25**, 329–341 (2019).
42. Alcolea, M. P. et al. Differentiation imbalance in single oesophageal progenitor cells causes clonal immortalization and field change. *Nat. Cell Biol.* **16**, 615–622 (2014).
43. Alcolea, M. P. & Jones, P. H. Cell competition: winning out by losing notch. *Cell Cycle* **14**, 9–17 (2015).
44. Lavker, R. M. & Sun, T. T. Heterogeneity in epidermal basal keratinocytes: morphological and functional correlations. *Science* **215**, 1239–1241 (1982).
45. Ghazizadeh, S. & Taichman, L. B. Organization of stem cells and their progeny in human epidermis. *J. Invest Dermatol* **124**, 367–372 (2005).
46. Silverman, B. W. Using kernel density estimates to investigate multimodality. *J. R. Stat. Soc.: Ser. B (Methodol.)* **43**, 97–99 (1981).

Acknowledgements

Author contributions

G.P. performed the computational simulations and carried out the analysis of experimental results. A.W. carried out the lineage tracing experiments. B.C. and D.F.A. performed H2BGFP-dilution experiments. A.H. and K.M. contributed with image acquisition and image analysis. V.K. performed preliminary, conceptual work and helped implementing the Non-Markovian simulation algorithm. B.A.H. and P.H.J. supervised and designed the work. G.P., B.A.H., and P.H.J. wrote the paper. All authors reviewed and edited the final version.

Competing interests

The authors declare no competing interests.

Additional information