Modelling mitosis with multiple phenotypes: relation to Haeckel's recapitulation law

Yuriy Alexandrov[1,2],

1. The Francis Crick Institute, 1 Midland Road, London, UK
2. Department of Physics, Prince Consort Road, Imperial College London, UK

Department of Physics, Imperial College London, London, SW7 2AZ, UK
E-mail address: yuriy.alexandrov@imperial.ac.uk
Photonics Group, Department of Physics, Prince Consort Road, Imperial College London, UK

Abstract

The article presents a novel stochastic mathematical model of mitosis in heterogeneous (multiple-phenotype), age-dependent cell populations. The developed computational techniques involve flexible use of differentiation tree diagrams. The applicability of the model is discussed in the context of the Haeckelian (biogenetic) paradigm. In particular, the article puts forward the conjecture of generality of Haeckel's recapitulation law. The conjecture is briefly collated against relevant scientific evidence and elaborated for the specific case of evolving/mutable cell phenotypes as considered by the model. The feasibility, basic regimes and the convenience of the model are tested on examples and experimental data, and the corresponding open source simulation software is described and demonstrated.

Keywords

Mitotic proliferation, phenotype, cell differentiation, Haeckel's law, stochastic modelling

MSC

92D25 Population dynamics, 92C15 Developmental biology, 92C30 Physiology (general)

1. Introduction

Explosive growth of information on molecular components and mechanisms of life constitutes one of the hallmarks of scientific modernity. The main contributors to this process are certainly the biochemistry and biomicroscopy R&D. This study however, only cherry picks some of the major findings of mainstream cell biology to devise the very minimalist computational model of cell proliferation dynamics. The approach is essentially phenomenological and complemented by a corresponding view of evolutionary theory.

The subject of mathematical modelling of cell phenotype dynamics is recurrent in the literature. One of the early but in-depth examples of nonlinear Ordinary Differential Equations (ODE) systems applications to immunology is the work of Perelson et al (1997),

which actually treats coupled rate equations for the mixture of different cellular and molecular components in blood. More recently, the ODE approach was applied, in particular, to model abnormal regimes in haematopoiesis with multiple phenotypes (Colijn et al 2005a, Colijn et al 2005b), and competition-mediated proliferation in two-phenotype cell culture (Bove et al 2017). It is well-known, however, that the right-hand sides of such ODEs are the subject of meticulous designs bordering with art, as ODEs, being a general method, are not specially suited for cell proliferation modelling. Among stochastic techniques it is worth mentioning the version of the two-phenotype stochastic model of concurrent haematopoiesis (Roeder et al 2002) extended for the case of several phenotypes (Glauche 2007) and the multi-phenotype model of bone marrow featured by cellular automaton-style considerations (Bessonov et al 2008), the latter being the closest to the presented approach. Another stochastic cellular automaton model (Sundstrom et al 2016) studied tumorigenesis with multiple phenotypes by taking into account 2D and 3D spatial effects and limitations, but not cell differentiation. The work (Lu et al 2007) analysed the mitosis in two-phenotype systems on the basis of the algorithm by Gillespie (1984), also without cell differentiation. In the work (Jones 2010), the Bayesian age-dependent branching process modelling was applied to simulate three-phenotype dynamics with differentiation; this method is conceptually quite complex but it does not work with an arbitrary lineage formation tree. An integrated PDE approach was used to devise the two-phenotype model of kidney cell proliferation in a mouse embryo to fit the Optical Projection Tomography (OPT) data (Lefevre et al 2013), and the two-phenotype model of hematopoiesis (Adimy et al 2005), but these techniques are also hard to generalise for the case of many phenotypes. The same is true for the diverse set of models reviewed in the work on colonic crypt and colorectal cancer modelling (Johnston 2008). The monograph (Kimmel et al) describes several modelling methods applicable to the problem (e.g. the Galton-Watson process); however, these techniques employ a high level of detail and mathematical complexity, likely excessive in many concrete contexts. The work (Stadler et al 2018) operates with phenotype-specific parameters similar to those employed in the current study but without focusing on the time dependencies of cell numbers.

Summarising this brief overview of computational methods, one can notice that neither of the existing modelling approaches develops methodology and software specifically for the mitotic multiple-phenotype systems which is both simple, intelligible and convenient for biologists as well as general and flexible enough to cover cross-physiology subjects. This issue is addressed in the modelling chapter, but first the article elaborates on the nature of the mentioned generality.

Namely, the article argues in favour of the abstract Haeckel law for the living (replication-based, evo-devo) systems. The “primordial rock” hypothesis is put forward to emphasize the common evolutionary origin of mitotic cell phenotypes and the general nature of what is vaguely recognized today as “epigenetic determinism” (Waggoner et al 2015), but likely can be attributed to the above mentioned Haeckelian recapitulation.

The next and major part of the study is then devoted to the stochastic, differentiation tree-based mathematical model of heterogeneous mitotic proliferation. The model involves only a handful of essential phenotype parameters, those being cell cycle duration, probabilities of differentiation and exit (which may include apoptosis), and the generation counting threshold. The last property might seem unusual but it actually isn’t. For example, an index identical to the generation counting threshold was introduced in the theoretical model (Jones 2010) describing brain cell generation; there, it was explained as “the number of cell cycles [the progenitor] cell must go through before it is competent to produce oligodendrocyte”. In the

modelling work on mitosis with cell differentiation (Hasenauer 2013), one of the presumptions was that "in most multicellular organisms the mother cell divides symmetrically into two daughter cells which inherit the age of the mother cell"; this "age knowledge" property of cells is also synonymic to a generation counting ability. The biochemical evidence that cells are capable of counting numbers of divisions and utilising this information in decision making, has been recently discovered (Li et al 2016, Bernitz et al 2016).

The article is concluded by examining properties and outcomes of the computational model. The explanatory/predictive capability of the model is confirmed on several test cases, including applications to real-world data on renal development in mice (Lefevre et al 2013) and proliferation of the MDCK cell line in culture (Bove et al 2017).

2. The Paradigm

As mentioned above, advances in microscopy and in the understanding of genetics in terms of molecular components became defining factors for studying cells at mesoscopic level. In corresponding considerations in the literature, the notions of a cell's functional "mission" and distinct phenotype are central. They reflect not only the existential meaning of physiological purpose, but also common origins and relational directionality - it isn't possible to get a phenotype from anything else except other (precursor, progenitor, predecessor, parent) phenotypes. It is worth mentioning that the term "phenotype" itself was used in the first place in order to describe physical traits and the morphological/functional sameness of macroscopic animals and plants (albeit later being agglomerations of distinct organs that are in their turn agglomerations of cells). In macroscopic (species)-level biology the term "phenotype" was associated with a whole package of notions dictated by evolutionary principles - mutations, feature heritability, intra- and interspecific competition, "selection pressure", as well as the well-known phylogenetic evolutionary tree diagrams. Another powerful albeit less frequently mentioned paradigm is Haeckel's law of embryonic development -" Ontogeny recapitulates Phylogeny"; its basis and meaning have remained a subject of intense theoretical debate for a significant period of time (e.g. (Laubichler et al 2009). Its modern formulation in zoology is the following: "The definite form of an animal develops on the top of the preexisting embryonic form of its ancestors" (Sharov 2014). Note that this formulation also allows recursion, as it implies some sort of directed (in general, one-to-many) relation between ancestral and descendant forms. The general formulation seems even simpler, being almost tautological: "The current form develops on the top of the prototype form (of its ancestor)". This is illustrated schematically in Figure 1. It is interesting to note that in recent literature, this "develop" keyword becomes more and more associated with the literal, "epigenetic" view of development as a consecutive, ordered activation (expression) of certain groups of genes (Waggoner et al 2015). The role of genes in recapitulation was also discussed in an earlier work by Ohno (1995). One can conclude, therefore, that any such concerted activation of genes may be (or rather cannot be anything else but) a part of, or an equivalent to "sliding along the branch" of a certain lineage tree of ancestral prototypes. This seems to agree well with the recent, eloquently articulated statement in the work (Stadler et al 2018): "macro-evolutionary processes happening over million of years may help us to understand microscopic processes on the level of cells".

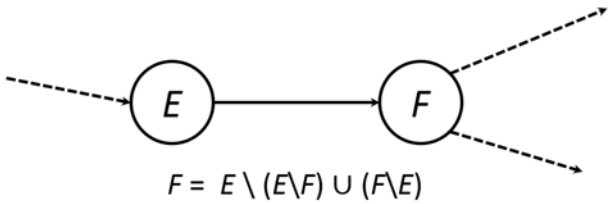

Figure 1 Formal representation of Haeckelian progression from form *E* to form *F*. The "form" (e.g. phenotype) is considered to be a set of well-defined sub-forms. The sets (*E**F*) and (*F**E*) denote some definite sub-forms developed in *E* but not in *F*, and vice versa. In other words, in order to progress to form *F* having form *E* as a base, one needs to lose some old and acquire some new sub-forms.

Despite many of the evo-devo ideas eventually becoming mainstream in the sub-cellular level literature, the relational tree diagrams for mammal cell phenotypes had not been studied in great detail until very recently (with the notable exception of haematopoiesis), and the cellular analogue to the Haeckel's recapitulation law is missing completely. It is interesting to note that in the work (Jones 2010), the mathematical formalism developed to model branching processes in molecular evolution (similar to the ones used in the BEAST (Bouckaert et al 2014) and MrBayes (Huelsenbeck et al 2001) programs) but also applicable for phylogenetic macroevolution, was adapted to describe progenitor cells' mitosis and differentiation in the brain, - but in this case the branching tree of cell types was not referred to as "phylogenetic".

The following statements, albeit hypothetical, are aimed at easing this disparity and hopefully bringing the subject of cellular recapitulation beyond semantics. They also try to elucidate the similarity between cells in terms of the common origin (and common place of origin) of cellular phenotypes.

1. Haeckel's recapitulation law is the general horizontal principle for all hierarchical levels of life driven by heritability and variation/selection (molecular, organelle, cell, organ, organism).
2. In particular, one can consider the well-known haematopoietic lineage tree based on hematopoietic stem cells (HSCs, Metcalf 2007) as "phylogenetic" in the Haeckelian sense. Haematopoiesis recapitulates the "embryonic" stages of evolution of mitotic pre-multicellular (and partly pre-eukaryotic) phenotypes in the succession of ancestral liquid organs that eventually became blood. This is the specific (albeit the most ancient) case of animal organogenesis.
3. The cellular "evolution" mentioned in (2) once took place in oceanic, shallow-water, calcium-rich geological structures ("primordial rock", a precursor to coral reefs). The environmental pressure caused by the reefs' different climatic niches has led to the accumulation of corresponding cell differentiation potentials in their genomes. The internal locus of the source of blood-forming and of other stem cells in the animal body (bone marrow, Sousa et al 2013) points at such origin directly. The similarity between the structures of bone and coral is a known fact (Bronzino 2006). It is also known that both calcium dependent bone-forming osteoblasts and bone-resorbing osteoclasts derive straight from bone marrow cells (Heinemann et al 2011). Yet another hint is calcium signalling, which is ubiquitous in the regulation of such a number of cellular functions, that it may be considered a defining feature of cellular life per se.

4. From this perspective, one can consider the “primordial broth” within that “primordial rock” as the ancestral organ of blood and lymph, and the rock itself as an ancestor of bone tissue.
5. Any cellular process involving proliferation, migration, differentiation and structure formation (e.g. embryo development, tissue regeneration, immune response, inflammation, cancer, bacterial infection, angiogenesis, etc.), “plays back” the corresponding pre-recorded elementary steps (patterns) of biotic evolution. In particular, biotic recapitulation is the only way of building a functional supra-cellular level biology (tissue/organ/organism). Certainly, the lineage-tree driven separation of proliferating phenotypes in time is in many cases followed by a concomitant separation in space; the well-confirmed HSC “niche” concept (Morrison et all, 2014) in that matter doesn’t differ from any other organ-specific stem cell pool or niche.
6. Cancer is a specific case of organogenesis, implementing pre-recorded phenotype progression routes that are, arguably, either archaic, or concomitant (parasitic) to mainstream lineage formation. In viable organisms, these biogenetic routes are either blocked or have negligible probability, but they may be initiated by mutation.
7. During development via Haeckelian recapitulation, cell populations are controlled by generation counting. Namely, the precursor phenotype cells count the number of divisions during mitosis and, after reaching some threshold number, progress to the (one of the) next allowed phenotype(s). It is reasonable to assume that such generation-counting control is behind organ scaling and allometry phenomena (Shingleton 2010). The evidence of as well as the prospective mechanisms and biochemical basis for cell division counting have recently been discovered (Li et al 2016, Bernitz et al 2016). In particular, in algal cell cultures, this counting mechanism is associated with a specific protein CDKG1 (Li et al 2016); the amount of this protein in the cell is reduced after each division.

3. Computational model

The computational recipes listed below do not essentially rely upon any of the potentially questionable aspects of the “paradigm” presented above, except for possibly the generation counting principle. However, the latter can be considered from a purely pragmatic viewpoint as well, by simply treating generation counting thresholds as adjustment parameters of the model. On the other hand, taking “The Paradigm” into account can be beneficial for further progress in understanding the evolutionary “logic” behind every concrete case of recapitulation-driven proliferation.

The main components and principles of the computational model are:

- Distinct cell phenotypes capable of differentiating from one to another.
- Stochasticity. This is essentially stochastic technique.
- Cell cycle. In the code, the corresponding parameters are the cell cycle duration T_c and its standard deviation ΔT_c ; they are different for different phenotypes.
- Differentiation lineage trees, implemented as directional graphs in the code, with every node representing a phenotype, and every edge representing the probability of differentiation from the current to the next allowed phenotype.
- Direct control of cell numbers by the exit probability.
- Generation-iterating method (“God’s algorithm”, see below).

- Generation-counting thresholds.

Among these principles, only the last two are novel, and as such require further specification.

The generation-iterating method is recruited as a shortcut to circumvent the difficulty of iterating the time-dependent state variables in stochastic modelling. If one chooses some fixed Δt as a time increment driving physiological dynamics, all results become dependent on the value of Δt. Using rate-dependent methods such as the Gillespie algorithm (Gillespie 1984) to overcome this adds significant mathematical and computational complexities. In the present study, the problem is avoided by using the observation, that in some cases, having a set of (g)-generation cells is fully sufficient for deducing the set of next (g+1) generation cells. In the simple implementation presented in this article, for every newly created (g+1)-generation cell, the following parameters are immediately defined:

1) If it exits the pool or not, instead of continuing (according to exit probability α)
2) The phenotype - either the parent's, or the next (depending on the generation count threshold)
3) Generation count - either the parent's+1, or 1 (ditto)
4) Birth time - equals the end of the life time of the parent cell
5) End of life time - (birth time + randomized T_c, depending on the phenotype)

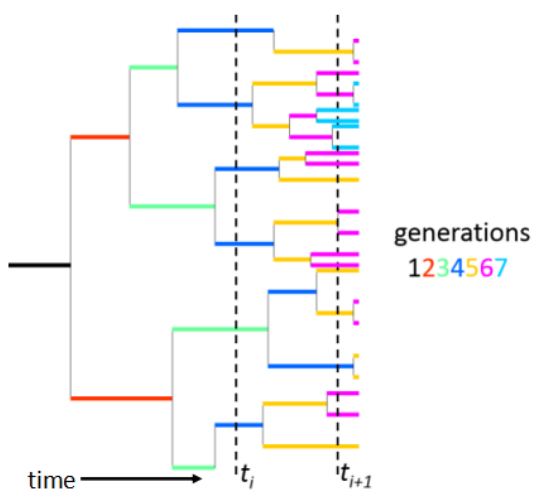

Figure 2 "God's algorithm" in action. Horizontal segments represent cells of different generations, correspondingly colour-coded. When computing the time dependencies of cell numbers, one needs to count the number of segment intersections ('hits') made by the sampling *t*-lines (dashed black vertical lines)

During simulations, properties 1, 2, 3 and 5 are stochastically chosen for every new cell, whereas properties 2, 3, 4 and 5 are actually kept in computer memory. The term "God's algorithm" reflects the usage of property 5, which implies "knowing the future" of a cell's fate at the time of its birth. Herewith, the cell lifetime values are randomized by sampling from a Gaussian distribution centred at T_c with standard deviation ΔT_c. Obviously, in order to estimate the number of cells, $N(t)$, as a function of time t from such simulations, one needs to count those cells for which time t falls within their lifespan (Figure 2). The drawback of the algorithm is that in order to do this, one needs to keep a complete record of the history of all cells in memory. Fortunately, this is not very problematic, since as was shown for the simplest case above, every cell is represented by only 4 numbers (type, generation and the

birth and end of life times). Its advantage is that it allows replacing iteration by time by the iteration of generations.

The influence of exit probability α on cell proliferation dynamics can be illustrated by the following considerations. Imagine N_0 cells of the same type, starting their divisions at time $t = t_0$ with average cell cycle duration T_c. If α has the meaning of the fraction of cells leaving the pool after each division, then the expected number of cells as a function of time, $N(t)$, can be written as

$$N(t) \approx N_0 \cdot [2(1-\alpha)]^{\frac{t-t_0}{T_c}} \qquad (1)$$

Formulae similar to (1) have been published in the literature (e.g. (Lefevre et al 2013, Tomlinson et al 1995). It is worth noting that in expression (1), the two essential parameters T_c and α are competing in the proliferation rate expression, since (1) obviously implies $N(t) \propto \exp\left(\left[\frac{\log(2(1-\alpha))}{T_c}\right]t\right)$. It is therefore important to have reasonable estimates of the cell cycle durations T_c when specifying the model. These can in many cases be provided by microscopy (e.g. Bove et al 2017, Lefevre et al 2013).

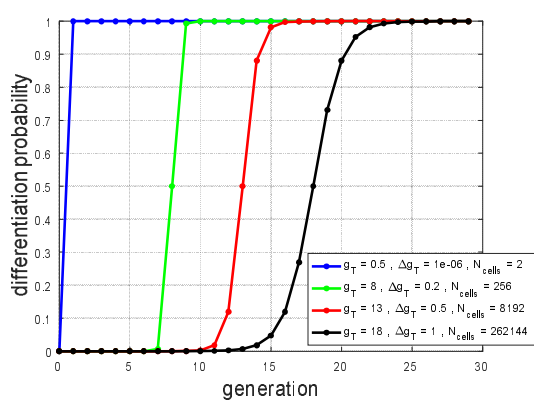

Figure 3 Sample phenotype progression probability dependencies as a function of local generation count, ranging from threshold-like transitions (blue) to smooth ones (black)

The last feature of the framework is the utilisation of generation counting thresholds. This is also straightforward. Certainly, having the proper number of intermediate proliferating phenotypes is critical, as their influence is amplified exponentially via mitosis. This is illustrated in Figure 3, where the corresponding cell numbers at threshold level are shown to be $N_{cells} = 2^{g_T}$, where g_T is the generation counting threshold (number of divisions the cells of this phenotype, on average, perform). In the implementation, not only the generation threshold g_T is taken into account, but also its dispersion Δg_T, which represents the degree of stochasticity in the model's decision-making. Both parameters are fed into a standard sigmoid expression to get the corresponding probability of progression to the next allowed phenotype. Thus, for a g-th generation cell, this differentiation probability $P_{diff}(g)$ reads as

$$P_{diff}(g) = \frac{1}{1 + \exp\left(-\frac{g - g_T}{\Delta g_T}\right)} \tag{2}$$

In simulations, the probability $P_{diff}(g)$ is used with a uniform, unit-interval distributed random variable to get a randomized "yes/no differentiation" response at the moment of every cell's birth.

Altogether, every phenotype in the model is characterized by 5 parameters (T_c, ΔT_c, g_T, Δg_T, and α), plus the out-degree (the number of outcoming edges) of the node in the directed lineage tree. Every edge in a tree is characterized by the weight representing phenotype-to-phenotype transition probability. For each node therefore, the weights of the outcoming edges sum up to 1. In the course of simulations, if the stochastic choice (made at the moment of cell birth with probability $P_{diff}(g)$ given by the equation (2)) favours cell differentiation, the cell differentiates into one of the types defined by the corresponding lineage tree edges outcoming from the parent node. Herewith, the next specific phenotype of a cell is defined by a roulette wheel method involving the weights of a preceding phenotype's outcoming edges.

It is quite obvious, that any change in the "flow of cell numbers" along the branches of the differentiation tree diagram will eventually lead to the formation of cell populations of differing size (as represented by final phenotypes), and therefore will cause effects such as organ scaling and allometry in embryo development. In the absence of exit, these "flows" of cell numbers from a given node are controlled by the weights of the outcoming edges and by the parameters of generation count thresholding in (2) responsible for the total number of cells of that node's phenotype.

Finalizing this section, it is relevant to note that the model allows reproducing biologically meaningful population dynamics with an arbitrary lineage tree without resorting to complex mathematics like the Bayesian inference. On the other hand, the model is neither limited by the specifically chosen proliferation method (iterating generations), nor by the absence of the dependence of cell phenotype parameters on the presence of cells of other types ("no competition" presumption).

The numerous and sophisticated models of cell interaction and competition have already been known for a relatively long time (e.g. Perelson et al 1997). Such improvements and extensions of the technique are possible and can be investigated in the future.

4. Model implementation

The model was coded in Matlab (Mathworks, Inc.). To represent and use differentiation lineage trees, ***digraph*** and related graph functions were used. A simple, intuitive GUI is provided. The computational performance of the model is satisfactory for simple cases involving about 10^4 cells. However, when the total number of cells exceeds 10^6, the algorithm slows down noticeably. No attempt has been made to improve performance via parallelising, although it seems possible.

The software is available at https://github.com/yalexand/CellPopSim.

5. Model properties, results and conclusions

Typical dynamic regimes can be examined with simple linear succession graphs, containing just 2 to 4 consecutive nodes. In doing so, the overall validity of the model is assessed, which is shown is the following. As to the basic properties, the model demonstrated the expected behaviour. Simulations usually start from a single initial proliferating cell of an initial phenotype. In the demonstrative examples drawn below, parameters T_c and ΔT_c are set for all phenotypes within a range of 15-25 ± 2h, which is typical for eukaryotes' mitosis.

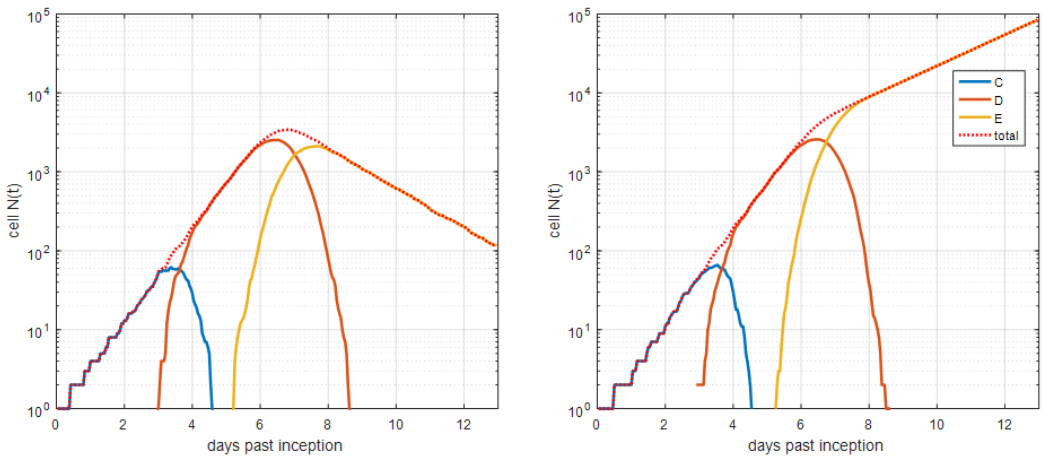

Figure 4 Sample linear three-phenotype succession $C \rightarrow D \rightarrow E$ showing two types of exponential dynamics for final phenotype E: "disappearance" at exit probability $\alpha=0.65$ (left), and unrestricted growth at $\alpha=0.35$ (right)

Figure 4 presents an example of a linear, three-phenotype succession showing two types of exponential dynamics for the final phenotype E: "disappearance" at exit probability $\alpha>1/2$ (Figure 4, left) and unrestricted growth at $\alpha<1/2$ (right).

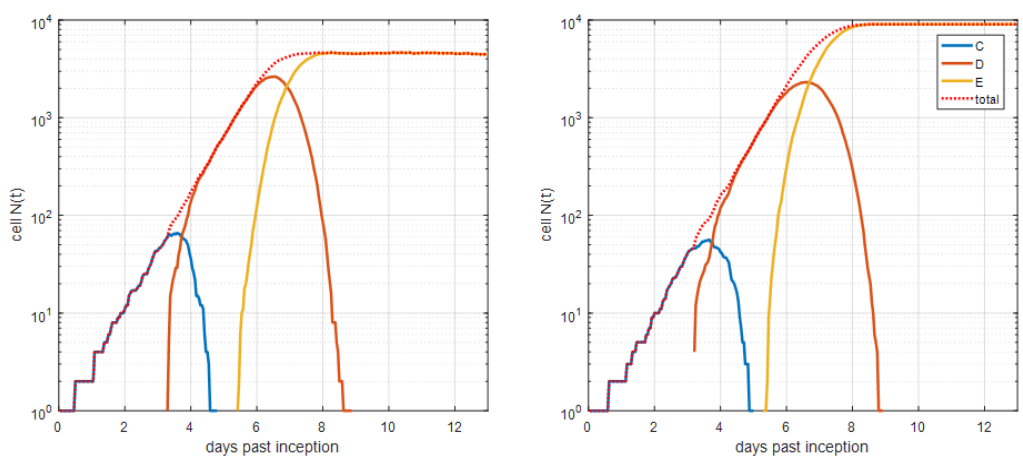

Figure 5 Sample linear three-phenotype succession $C \rightarrow D \rightarrow E$ showing the two types of "stable state" for final phenotype E: "dynamic" at exit probability $\alpha=0.5$ (left), and genuine quiescence at $\alpha=0$, $T_c=\infty$ (right)

The Figure 5 compares two ways of reaching equilibrium for the final phenotype. In the first, mitosis continues indefinitely with α=1/2 but also with g_T=∞ (exit-mediated equilibrium). The second is establishing genuine quiescence at α=0, T_c=∞, in which case g_T becomes irrelevant (Figure 5 left and right, correspondingly). The higher level of the plateau on the right plot is due to the α=0 setting, which preserves all created *E*-cells.

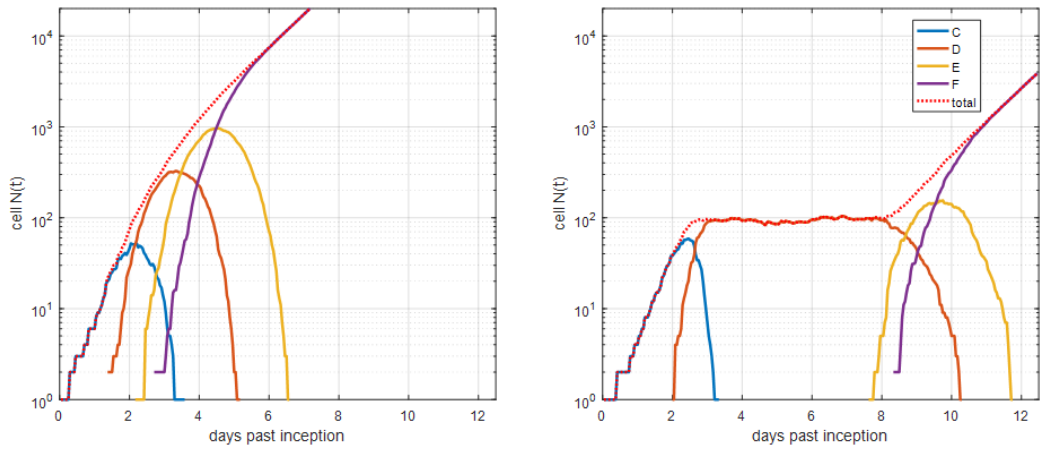

Figure 6 Sample linear four-phenotype succession *C*→*D*→*E*→*F* showing unhindered exponential growth (left), and delayed "α=1/2 carrier"-mediated growth (right)

The next example presents the four-phenotype linear succession which keeps overall exponential growth (Figure 6, left), and the same system with a higher (g_T=10 vs. 3 for the plot on the right) generation count threshold and an α=1/2 setting for the phenotype *D* (vs. α=0 for the plot on the right). The α=1/2 setting removes exactly half of the newly born phenotype *D* cells during proliferation. Altogether these settings for phenotype *D* allow delaying the exponential growth of phenotype *F* for about 4 days (Figure 6, right). Therefore, such an intermediate phenotype with α=1/2 and higher g_T can serve as a "carrier to the future" for subsequent phenotypes, providing the needed flexibility, e.g. in the case of delayed organogenesis.

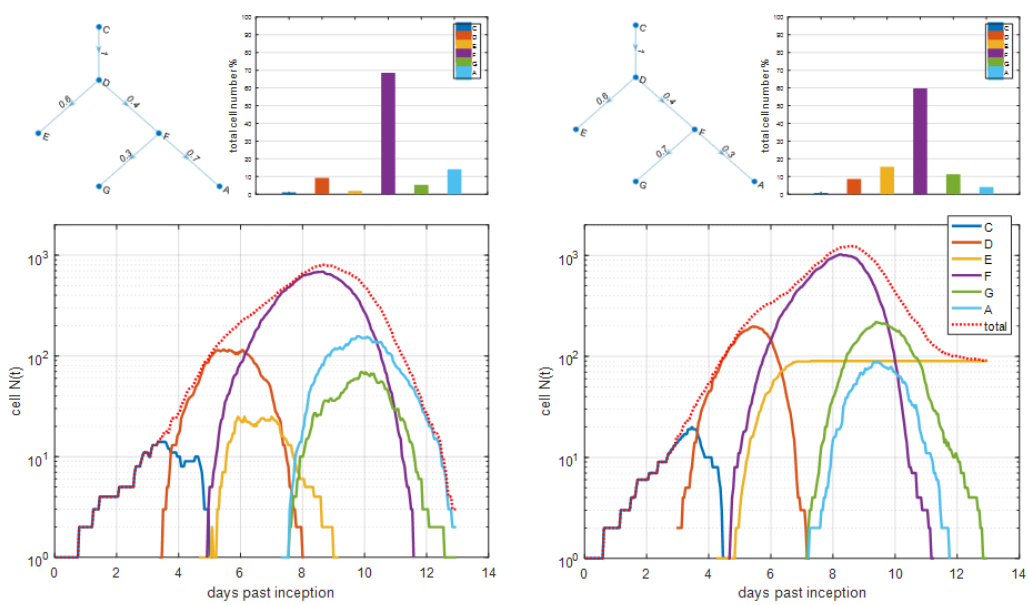

Figure 7 Sample forked five-phenotype proliferation "outbursts". The terminating nodes *E*, *G* and *A* are all subjected to excitation probability α=0.8, which provides fast decay (left). On the graph to the right, the differentiation probabilities for nodes *G* and *A* (edges *F*→*G* and *F*→*A*) are swapped, resulting in their curves changing places, whereas the *E* node was made the quiescent "single survivor" at $T_c = \infty$

The simple, six-phenotype forked proliferation outburst? is shown in Figure 7. On the graph to the left, the terminating nodes (final phenotypes) *E*, *G* and *A* are all "dumped" to decay by setting α=0.8. On the graph to the right, the differentiation probabilities for nodes *G* and *A* (edges *F*→*G* and *F*→*A*) are swapped, resulting in their curves changing places, whereas the *E* node was made the quiescent "single survivor" by the setting $T_c = \infty$. The bar plots display the relative number of cells of different types in the studied time window, which makes sense for the case of "outburst".

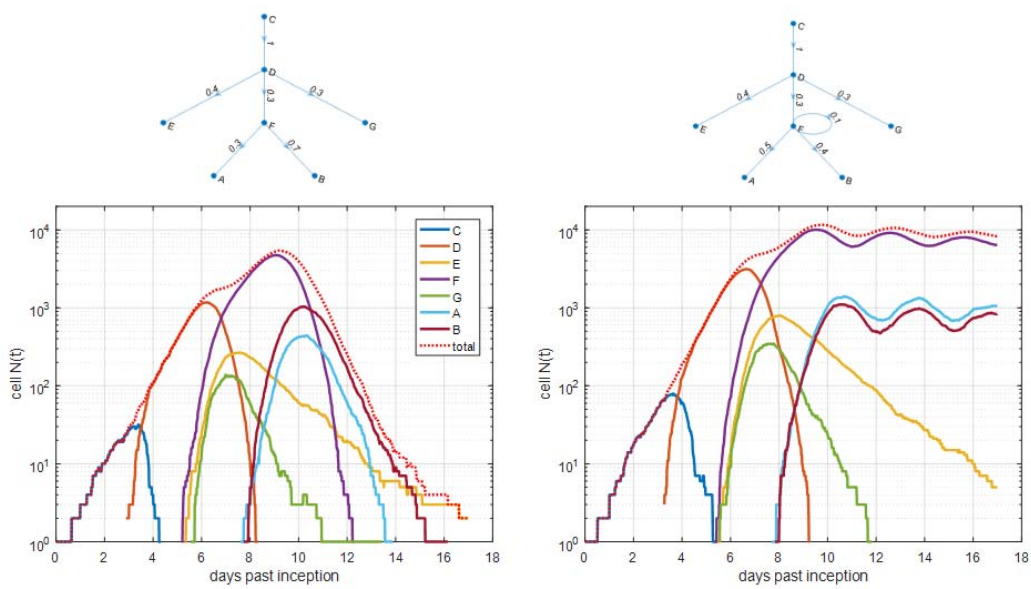

Figure 8 Sample forked seven-phenotype proliferation progressing from "outbursts" (left) to undulations (right) after the addition of a self-renewal property to node *F*. Similarly to the example shown in Figure 6, terminating nodes *E*, *G*, *A* and *B* are all subjected to excitation probability α=0.8

Another example presents the phenomenon of undulations initiated in the seven-phenotype "outburst"-type branching system (Figure 8, left) when one of the branching nodes is assigned a self-renewing property whereas all other parameters stay the same (Figure 8, right). The undulations develop due to the interplay between cell cycle duration T_c and the generation-counting threshold g_T, which demands the periodic reset of the generation counter to g=1 for the self-renewing phenotype. Certainly, if g_T<1 and $\Delta g_T << g_T$, the only two generations allowed for this phenotype are either g=1 or g=2; this is the way to effectively "undo" generation counting. It is hard to say if a self-renewing node with g_T >1 (as in Figure 8, right) can represent any biological reality. Nevertheless, as this explanation shows, such hypothetical effects of "self-renewal undulations" at g_T >1 are intrinsic to the model. On the

other hand, the involvement of many unrelated proliferating sources will cause undulation smearing, in which case undulations will not be noticeable in the experimental data.

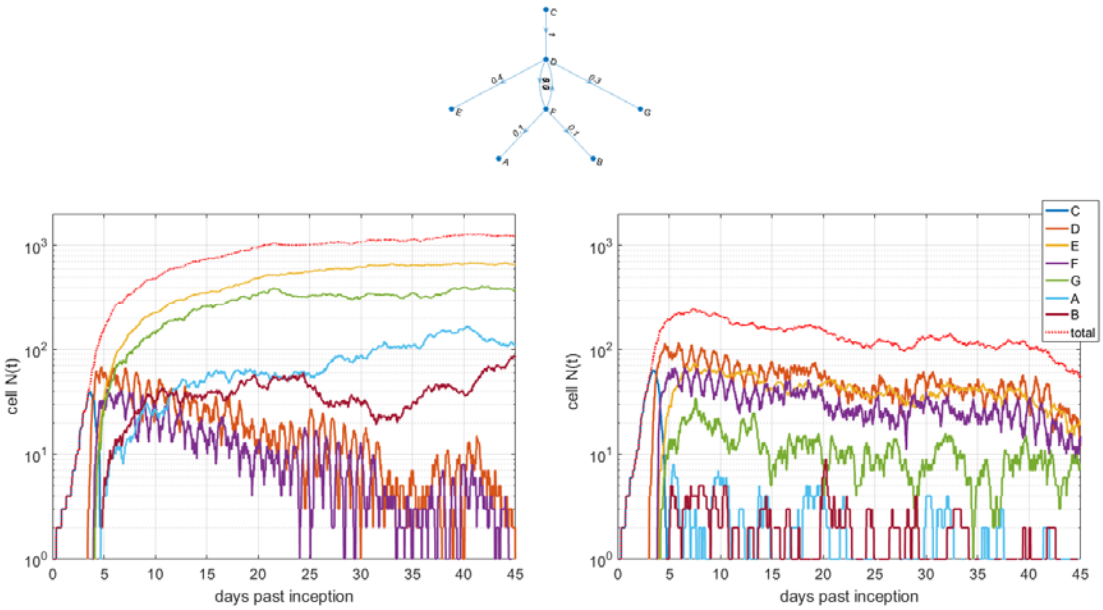

Figure 9 Sample forked seven-phenotype proliferation progressing from a "stable state" where the terminating nodes *E*, *G*, *A* and *B* are all subjected to excitation probability α=0.5 (left) to decaying "outbursts" at α=0.8 (right). In both cases, the undulations in *F* and *D* cell numbers are caused by de-differentiation introduced by edge $F \rightarrow D$

Another example of undulations, this time caused by introducing the de-differentiation edge, is shown in Figure 9. In this case, the generation counting was "undone" for both involved nodes as described above, by setting g_T<1 and $\Delta g_T << g_T$, so that the period of such undulations is defined only by the cell cycle duration values, T_c, of these nodes. Undulations also develop when terminating nodes are involved (data not shown). It is relevant to note that the undulations of cell numbers have been reported in haematopoiesis data and studied in corresponding models (Colijn et al 2005a, Colijn et al 2005b).

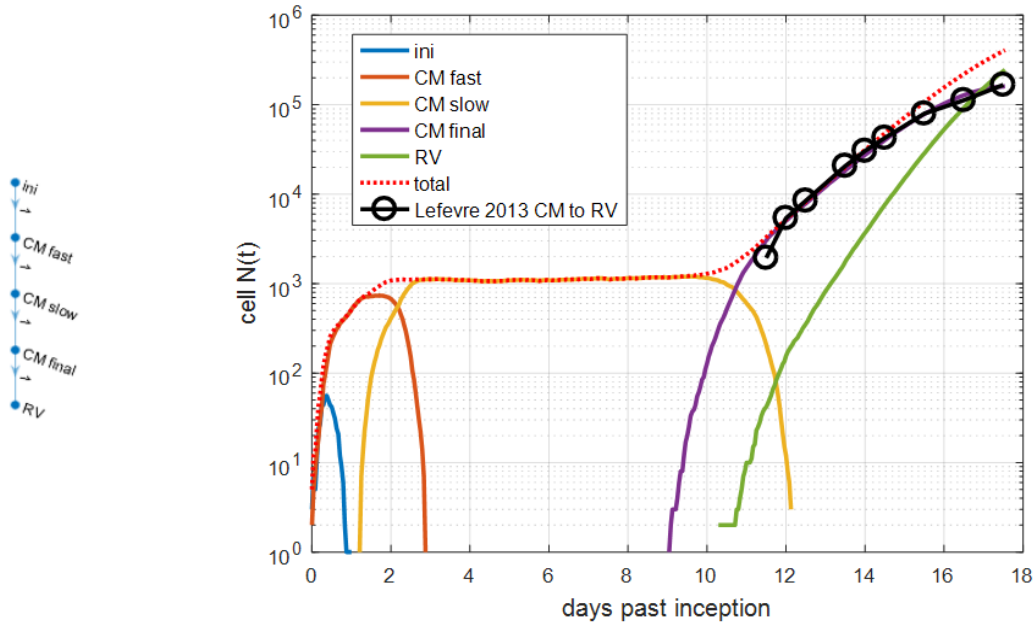

Figure 10 Application of the model to renal organogenesis by fitting OPT microscopy data published by Lefevre et al (2013). The data are used with the authors' permission. Full details in Section 5

The next example (Figure 10) represents an attempt to fit the experimental OPT data reported in the work on mouse kidney development by Lefevre et al (2013). This study reports a very steep near-exponential growth of the mass of the precursor cap mesenchyme (CM) cells at about 2 weeks past conception, before differentiation to the final renal ventricular (RV) cell type. In the article, the volume of the CM cells was reported in voxels, with 1 voxel = 4.17μm^3. Assuming a cell radius of a≈6μm, the estimate for cell volume is $(4/3\pi a^3)/4.17\approx217$[vox], and therefore Data[vox]/217 should be close to CM cell numbers. The first "ini" phenotype is used as an adjustment. The article reports two CM sub-phenotypes at around 2dpc with T_c =19.6h and T_c =15h and comparable amounts, which was implemented in the model. The "slow CM" is then used as a "carrier" precursor to the point where the fast growth of CM cells starts ("final CM"). The resulting adjustment for this phenotype was done for the "no apoptosis" case, α=0, as suggested in the article, and T_c =18h. The RV phenotype's settings included α=0 and T_c =∞. As shown in Figure 10, the model allows presenting the appearance of the final RV phenotype at a very early stage of renal development and in a quantitative manner. The more rigorous application of the technique can involve parameter estimation and uncertainty analysis, as well as larger experimental datasets (as compared with the number of model parameters); however, this is beyond the scope of the current study.

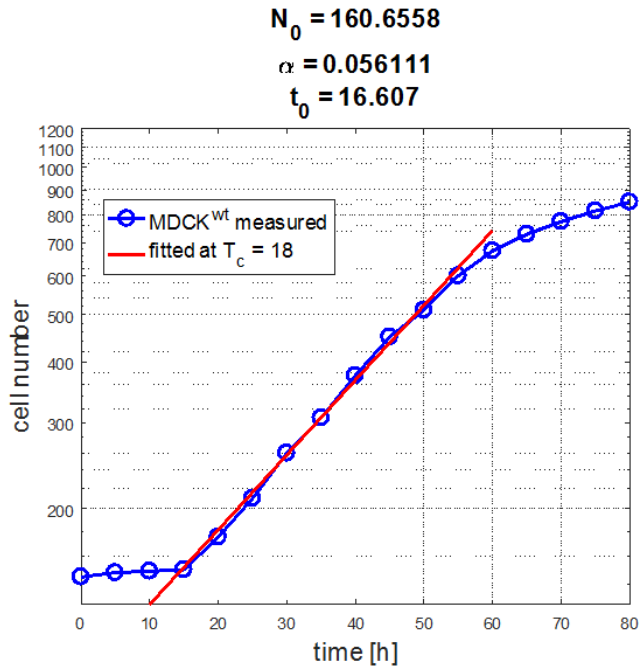

Figure 11 Fitting of the time dependence of the number of proliferating MDCK cells published in Bove 2017 to expression (1), performed using the Matlab *fminsearch* function. Note the log Y scale. The data in Fig 11, 12 are used with the authors' permission

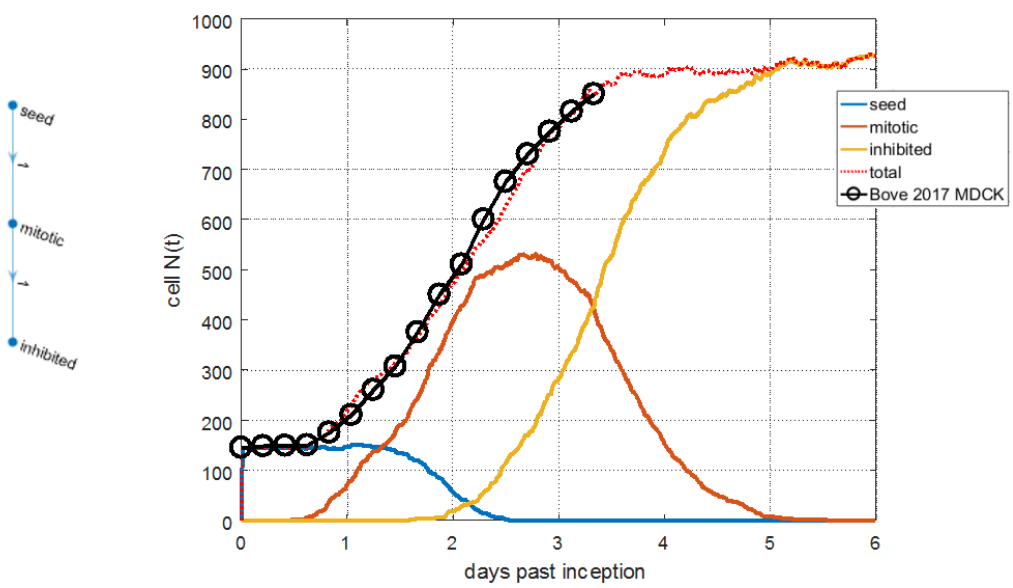

Figure 12 Adjustment of the time dependence of the number of proliferating MDCK cells published in Bove 2017 using a three-stage proliferation model

The last example considered in the study is the volume-limited proliferation of the immortalized MDCK cell line in a Petri dish, reported in a study of homeostasis with cell competition (Bove et al 2017). The competing phenotype was represented by the same MDCK cells, but lacking the polarity protein scribble. The present study, however, focuses on the basic (wild type) $MDCK^{wt}$ cells only (the green curve in Figure 2f of the article). The $MDCK^{wt}$ cell numbers, as a function of time (Figures 11 and 12) and cell cycle duration T_c

=18h were measured with very high precision. As shown in the Figure 11, first, the approximately exponential middle part of the dependency of cell number on time is chosen in order to fit it to expression (1); fitting was performed with the cell cycle duration fixed at T_c =18h as suggested in (Bove et al 2017). Fitting was carried out by a multidimensional nonlinear minimization (Nelder-Mead, MATLAB "fminsearch" function). This allowed estimating the apoptosis probability, $\alpha \approx 0.056$, and the starting cell number, N_0. Next, these parameters were used to define the "seed" and the middle "mitotic" phenotype in the simple chain proliferation model with three "phenotypes" termed "seed" (proliferating, with $\alpha=0$ and $g_T<1$), "mitotic" (proliferating with T_c =18h and $\alpha=0.056$) and "inhibited", respectively (Figure 12). It turns out, that in order to find a good adjustment for the whole dataset, it is enough to set the apoptosis rate of the "inhibited" phenotype to the familiar value of $\alpha=½$. In the experiment and the article's discussion, these three "phenotypes" were certainly not distinguishable from one another. However, this example, as well as the previous one, demonstrate that such decomposition of cells into groups according to their "mission" can be applied in a variety of cases to describe hypothetical population dynamics. In particular, it allowed interpreting volume-controlled proliferation with the help of a new, "inhibited" phenotype.

In conclusion, it is worth emphasizing again that the motivation and intrigue behind the developed computational model of cell proliferation was the conjecture of the generality of Haeckel's recapitulation law as propounded in "The Paradigm" section. According to this conjecture, Haeckel's law is one of the overarching principles of biological causality. The presented model shows how it can contribute to the understanding of mitotic proliferation with multiple differentiating phenotypes. Although the exact molecular mechanisms of genetics-driven recapitulation have yet to be discovered in full, it is believed that with the presented approach, coupled with the support of viable mathematical models, mechanistic "recapitulation physiology" can be introduced and further utilised to the benefit of biological R&D as a whole.

6. Acknowledgments

The author thanks Mr. Oleg Alexandrov for revising the English of the manuscript. Author gratefully acknowledges partial funding from the UK Medical Research Council (MRC, MR/K015834/1), the Biotechnology and Biological Sciences Research Council (BBSRC, BB/M006786/1).

7. References

Adimy M, Crauste F, Ruan S (2005) A Mathematical Study of the Hematopoiesis Process with Applications to Chronic Myelogenous Leukemia, SIAM Journal on Applied Mathematics, 2005, Vol. 65, No. 4 : pp. 1328-1352

Bernitz JM, Kim HS, MacArthur B, Sieburg H, Moore K, (2016) Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions, Cell. 2016 Nov 17;167(5):1296-1309.e10. DOI: 10.1016/j.cell.2016.10.022. Epub 2016 Nov 10

Bessonov N, Pujo-Menjouet L, Volpert V (2008) Cell Modelling of Hematopoiesis, Math. Model. Nat. Phenom, Vol. 2, No. 1, 2008, pp. 81-103

Bove A, Gradeci D, Fujita Y, Banerjee S, Charras G, Lowe AR (2017) Local cellular neighborhood controls proliferation in cell competition, Mol Biol Cell. Nov 7;28(23):3215-3228. DOI: 10.1091/mbc.E17-06-0368. Epub 2017 Sep 20

Bouckaert R, Heled J, Kühnert D, Vaughan, Wu C-H, Xie D, Suchard MA, Rambaut A & Drummond AJ (2014) BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput Biol 10, e1003537.

Bronzino J (2006) ed, Biomedical Engineering Fundamentals, 39.3.3 39.13, ed Joseph Bronzino 2006, Taylor and Francis group LLC, ISBN-13:978-1-4200-0285-7

Colijn C, Mackey M (2005) a A mathematical model of hematopoiesis: I. Periodic chronic myelogenous leukaemia, Journal of Theoretical Biology 237(2), 117-32

Colijn C, Mackey M (2005) b A mathematical model of hematopoiesis: II. Cyclical neutropenia., Journal of Theoretical Biology 237(2), 133–146

Gillespie D (1984) Markov Processes: An Introduction for Physical Scientists. San. Diego: Academic Press Inc., New York 1984.

Glauche I, Gross M, Loeffler M, Roeder I (2007), Lineage Specification of Hematopoietic Stem Cells: Mathematical Modeling and Biological Implications, STEM CELLS 2007;25:1791–1799

Jones G (2010) Calculations for multi-type age-dependent binary branching processes, Journal of Mathematical Biology, 2010, DOI: 10.1007/s00285-010-0362-8.

Jonston M D (2008), Mathematical modelling of cell population dynamics in the colonic crypt with application to colorectal cancer, PhD Thesis, Oxford University, 2008

Hasenauer J (2013), Modelling and parameter estimation for heterogeneous cell populations, Thesis, Stuttgart University 2013, p.59

Heinemann C, Heinemann S, Worch H Hanke T (2011) DEVELOPMENT OF AN OSTEOBLAST/OSTEOCLAST CO-CULTURE DERIVED BY HUMAN BONE MARROW STROMAL CELLS AND HUMAN MONOCYTES FOR BIOMATERIALS TESTING, Eur Cell Mater Jan 25;21:80-93, DOI: 10.22203/eCM.v021a07

Huelsenbeck J P, Ronquist, F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics, 17, 754-755. doi:10.1093/bioinformatics/17.8.754.

Kemmel et al 2002, Marek Kimmel, David E. Axelrod, Branching Processes in Biology, 2002, Springer-Verlag New York Inc.

Laubichler M D, Maienschein J (2009) From Embryology to Evo-Devo, ed.:

Series: Dibner Institute Studies in the History of Science and Technology, The MIT Press (2009), ISBN-10: 026251334X, ISBN-13: 978-0262513340

Lefevre J, Marshall DJ, Combes AN, Ju AL, Little MH, Hamilton NA (2013) Modelling cell turnover in a complex tissue during development, J Theor Biol. 2013 Dec 7;338:66-79. DOI: 10.1016/j.jtbi.2013.08.033. Epub 2013 Sep 7

Li Y, Liu D, López-Paz C, Olson BJ, Umen JG (2016), A new class of cyclin dependent kinase in Chlamydomonas is required for coupling cell size to cell division, Elife. 2016 Mar 25;5:e10767,,PMID: 27015111 PMCID:PMC4841777, DOI:10.7554/eLife.10767

Lu T, Shen T, Bennett M, Wolynes P, Hasty J (2007) Phenotypic variability of growing cellular populations. Proc Natl Acad Sci USA 104: 18982–18987

MATLAB minsearch function, (https://mathworks.com/help/matlab/ref/fminsearch.html)

Metcalf D, On Hematopoietic Stem Cell Fate, Immunity Commentaries, DOI 10.1016/j.immuni.2007.05.012

Morrison SJ, Scadden DT (2014), The bone marrow niche for haematopoietic stem cells. Nature. 2014; 505(7483):327–34. doi:10.1038/nature12984.

Ohno S (1995) Why ontogeny recapitulates phylogeny, Electrophoresis. Sep; 16(9):1782-6.

Perelson AS, Weisbuch G (1997) Immunology for physicists, Rev. Mod. Phys. 69, 1219 – 1267, Published 1 October 1997

Roeder I, Loeffler M (2002) A novel dynamic model of hematopoietic stem cell organization based on the concept of within-tissue plasticity, Exp Hematol. 2002 Aug;30(8):853-61

Sharov AA (2014) Evolutionary constraints or opportunities?, Biosystems. ; 0: 21–30. DOI:10.1016/j.biosystems.2014.04.005.

Shingleton A (2010) Allometry: the study of biological scaling, Nature Education Knowledge 3 (10), 2, 2010. 61, 2010.

Sousa BR et al (2013) Human adult stem cells from diverse origins: An overview from multiparametric immunophenotyping to clinical applications, Cytometry Part A, DOI 10.1002/cyto.a.22402

Stadler et al (2018), On the statistical analysis of single cell lineage trees. T Stadler, S Skylaki, KD Kokkaliaris, T Schroeder. Journal of theoretical biology 439, 160-165, 2018.

Sundstrom A, Bar-Sagi D, Mishra B (2016) Simulating Heterogeneous Tumor Cell Populations.PLoS One. 2016 Dec 28;11(12):e0168984. doi: 10.1371/journal.pone.0168984. eCollection 2016.

Tomlinson I, Bodmer W (1995) Failure of programmed cell death and differentiation as causes of tumors: some simple mathematical models, Proceedings of the National Academy of Sciences U.S.A., 92,11130-11134, 1995

Waggoner M, Uller T (2015) Epigenetic Determinism in Science and Society, New Genet Soc. Apr 3; 34(2): 177–195