Some lessons and perspectives for applications of stochastic models in biological and cancer research

Alan U. Sabino[1,2], Miguel F. S. Vasconcelos[1,2], Misaki Y. Sittoni[1,2], Willian W. Lautenschläger[1,2],
Alexandre S. Queiroga[1,2], Mauro C. C. de Morais[1,2], and Alexandre F. Ramos[1,2*]

[1] *Escola de Artes, Ciências e Humanidades & Núcleo de Estudos Interdisciplinares em Sistemas Complexos,*
Universidade de São Paulo, Av. Arlindo Béttio,
1000 CEP 03828-000, São Paulo, SP, Brazil and
[2] *Departamento de Radiologia – Faculdade de Medicina,*
Universidade de São Paulo & Instituto do Câncer do Estado de São Paulo, São Paulo, SP CEP 05403-911

Randomness is an unavoidable feature of inner cellular environment and its effects propagate to higher levels of living matter organization such as cells, tissues, and organisms. Approaching those systems experimentally to understand their dynamics is a complex task because of the plethora of compounds interacting in a web that combines intra and inter level elements such that a coordinate behavior come up. Such a characteristic points to the necessity of establishing principles that help on the description, categorization, classification, and the prediction of the behavior of biological systems. The theoretical machinery already available, or the ones to be discovered motivated by biological problems, can play an important role on that quest. Here we exemplify the applicability of theoretical tools by discussing some biological problems that we have approached mathematically: fluctuations in gene expression and cell proliferation in the context of loss of contact inhibition. We discuss the methods that we have employed aiming to provide the reader with a phenomenological, biologically motivated, perspective of the use of theoretical methods. Furthermore, we discuss some of our conclusions after employing our approach and some research perspectives that they motivate.

Keywords: Markov Chains; Models, Theoretical; Stochastic processes; Regulation, gene expression; Contact Inhibition

I. INTRODUCTION

Description of Physical phenomena by means of theoretical methods has motivated building a rich machinery spreading from general relativity (description of matter behavior at macroscopic scale) to quantum mechanics (description of matter behavior at microscopic scale), from electromagnetism (description of electric charges, magnetic dipoles and light related phenomena) to condensed matter theory (microscopic description of solid state systems). Those tools have enabled the capability of controlling and designing specific experiments which outcomes are, in general, predicted within specific error ranges or to develop new technologies derived from that knowledge. Fortunately, theories have a scope of applicability which means they do not explain all observed data related to a given phenomenon. In general, that motivates elaborating new theories that may raise additional predictions, *i.e.*, additional verifiable hypothesis. For example, differently of Newtoniang gravity, general relativity succesfully predicts precession of the perihelion of Mercury or light bending by Sun. Furthermore, experiments aiming to investigate different manifestations of a phenomenon would require the development of specific theoretical or technological tools. For instance, one may consider the use of tensor calculus in general relativity instead of vector calculus of Newtonian gravitation, or the high precision instruments required for detection of grav-

itational waves. Biology, on the other hand, has followed a different historical trajectory with a predominant use of experimental methods. Biologists also rely on qualitative models to aid on the construction of a static picture of biological phenomena. Such an approach has relevant scientific and technologic implications. For example, one may consider the establishment of evolution theory – a key paradigm of modern science – or the capability of controlling biological phenomena at molecular level – as it is the case of human insulin production. However that strategy has a clear limit if one is interested on the dynamics resulting of the interaction of a plethora of compounds at different levels of living matter organization, such as organisms, tissues, cells and molecules. Additionally, interactions among elements of different levels give birth to a highly complex picture which description will demand the use of all machinery available in the scientific toolbox. That includes the use of mathematical methods not only as a crunching number technique but also as a strategy for formulating new principles that might be useful on description of biological phenomena, on testing hypothesis not accessible experimentally, and, for the case of successful theories, predicting the outcomes of different experimental designs, or guide the development of new technologies.

In this mini review we approach the biological usefulness of quantitative techniques in investigation of biological phenomena. We consider the application of stochastic methods to describe phenomena happening at molecular and cellular levels. The first topic will be reviewed within the scope of a stochastic model for binary regulation of expression of a gene that is either self-repressed

* alex.ramos@usp.br

or externally regulated. The second topic is approached in terms of a stochastic model aiming to quantify the role of contact inhibition in a co-culture *in vitro* experiment combining keratinocytes and melanoma cells. Our molecular level investigations have enabled us to understand the characteristics of chemical kinetics of ON and OFF switch of a gene that leads self-repression to be the mechanisms of noise reduction on amounts of expressed proteins of a given gene [1]. Furthermore, we used this model to approach noise in development of *D. melanogaster* embryos. That helped us to understand how the externally regulating gene enables mRNA production and spatio-temporal patterning with proper precision during development [2]. At cellular level, it was shown that cell proliferation mechanism denoted as contact inhibition can be quantified as an exclusion diameter between cells. That mechanism enables the formation of clusters of melanoma in co-culture with keratinocytes [3]. Furhtermore, that model predicts that melanoma cells will prevail in a given spatial domain if one watches the cell population dynamics for a sufficiently long interval because of its lower degree of contact inhibition (or smaller exclusion diameters).

The intrinsic randomness of biological phenomena justifies the use of a stochastic approach for their investigation. At intracellular level randomness is caused by chemical reactants being present in low copy numbers and their heterogeneous distribution inside the cell [4]. For example, random fluctuations have been widely observed in gene expression of either prokaryotic and eukaryotic cells by means of fluorescence techniques [5–20]. Despite randomness inevitability different gene regulatory strategies may give raise to different noise behaviors. For example, a self-repressing gene will present a higher precision on controlling the amounts of its products [1, 21–24]. Alternatively, external regulation has been noticed as a gene regulatory strategy resulting in higher noise [5, 25, 26]. Those results suggest self-repression as the unique mechanism controlling gene expression when high precision is necessary, as it is the case during development. However, recent results show that external regulation may be sufficient to generate the required spatial precision for gene products stripes formation along the anterior-posterior axis of a *D. melanogaster* embryo [2, 27, 28].

Indeed, developmental processes require a high precision on the control of the amounts of specific gene products such that they are present at proper positions and times during the life of an organism. That fact may induce a perception that noise is always detrimental to the cell. Such a premise is not always true. For instance, individual cells increase their survival chances under stress conditions by means of noise in gene expression with the consequential generation of phenotype diversification [29–32]. Normally behaving tissues are characterized by well organized cellular structures along space and time. That achieved by means of homeostatic mechanisms controlling cell densities in tissues. However, molecular fluc-

tuations may affect cell genetics, induce uncontrolled expression of cell proliferation related genes, and induce the appearance of carcinoma *in situ*. The latter generates spatially less organized cell structures in tissues, breaks homeostatic behavior, and provides conditions for invasive cell phenotype to appear. That is a manifestation of stochasticity being a beneficial trait for cancer cells (at individual level) even though at an organismic level the noise has lethal effects after a while.

Therefore, an important challenge for cancer biologists is to determine mechanisms underpinning the progression of a carcinoma *in situ*, and how those cells become prevalent within a region for a sufficiently long interval such that invasive phenotype start appearing. One important mechanism necessary for the prevalence of the tumor cells is loss of contact inhibition [33, 34]. Contact inhibition of proliferation in culture experiments is associated with the ability of cells to maintain their density in a given tissue at optimal values [35, 36]. That causes cancer cells to keep proliferating in culture experiments even when confluence is reached [33]. In contrast, it has been shown that hypersensitivity to contact inhibition in fibroblasts of naked mole-rats is a mechanism that stops proliferation at lower cell densities in culture experiments. That is caused by the interplay between p16 and p27 cyclin-dependent kinase inhibitors which stops proliferation at lower densities than when p16 gene is not expressing such that cell densities reach that level observed in mouse [37]. Those experimental results suggest the necessity of finding a quantitative description of the intensity of contact inhibition in normal or cancer cells to enhance our capability of predicting or describing carcinoma *in situ* growth.

The next sections are devoted to give an overview of the three applications discussed above and to present some research perspectives. We start discussing the stochastic model for regulation of gene expression. We discuss the chemical kinetics that enable the self-repressing gene to be expressed at lower noise regimes. Furthermore, we present our results in the context of development of *D. melanogaster* embryos which indicates the possibility of using of this model to approach complex organisms. Then we discuss our cell level approach to quantify the degree of contact inhibition between two cells as an exclusion diameter. Lower degrees of contact inhibition are quantified by smaller exclusion diameters and that is applied on the description of a co-culture experiment with interacting melanoma and keratinocytes. We present some possibilities for future investigations at last section.

II. RANDOM FLUCTUATIONS IN GENE EXPRESSION

Randomness in gene expression has been measured in terms of the Fano factor, denoted by $\mathcal{F}$, which is defined as the ratio of the variance to the average. We denote the

number of gene products by n (the number of proteins or mRNA's) and the Fano factor is written as:

$$\mathcal{F} = \frac{\langle n^2 \rangle - \langle n \rangle^2}{\langle n \rangle}. \qquad (1)$$

The Fano factor provides a measure to compare a probability distribution with the Poissonian distribution. The Poissonian distribution has $\mathcal{F} = 1$ while $\mathcal{F} < 1$ characterizes a sub-Poissonian distribution (a distribution that is thinner than the Poissonian). The a super-Poissonian distribution has $\mathcal{F} > 1$ and is more spreaded than the Poissonian. Determining the probability distribution governing a the gene products number is important because it provides some hints on the regulatory strategy of the gene. For example, a constitutive gene will have n being a random variable governed by a Poisson distribution, a sub-Poisson distribution governs the self-repressing gene [1], while super-Poissonian distributions might indicate a positive feedback (governed by a bimodal distribution) or bursty expression (governed by gamma or negative binomial distribution).

The above analysis is completed in the scope of regulation of gene expression modeled by means of a binary promoter that is ON or OFF. When the promoter is ON there is synthesis of gene products at rate k while no synthesis happens in the OFF state. The gene products decay at a rate ρ. The rate of promoter switching from OFF to ON state is denoted by f while the opposite transition happens at rate h_1 (for the self-repressing gene) or rate h_2 (for the externally regulating gene). Fig. 1 presents a cartoon of our simplified model for regulation of gene expression.

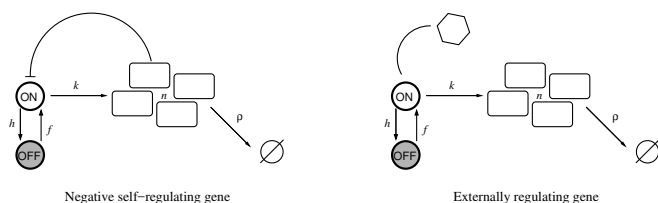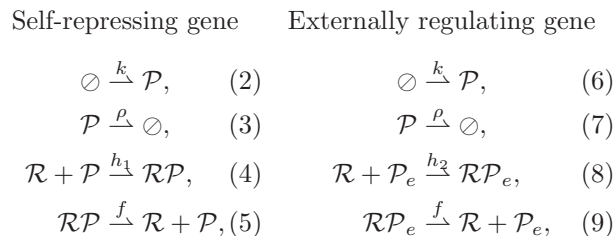Negative self–regulating gene   Externally regulating gene

FIG. 1: The left cartoon is a representation of a self-repressing gene while the right cartoon represents an externally regulated gene.

The scheme presented in Fig. (1) correspond to the set of effective chemical reactions presented below. The left hand side equations correspond to the self-repressing gene while externally regulated gene effective reactions are presented on the right. For the SRG we denote a protein by $\mathcal{P}$. The regulatory region of the gene is denoted by $\mathcal{R}$ and the gene state is determined by the binding of $\mathcal{P}$ to the regulatory region. The regulatory protein of the externally regulated gene is denoted by $\mathcal{P}_e$. The symbols for reaction rates appear on top of arrows indicating the reactants and products of the effective reactions.

Self-repressing gene   Externally regulating gene

$$\oslash \xrightarrow{k} \mathcal{P}, \quad (2) \qquad\qquad \oslash \xrightarrow{k} \mathcal{P}, \quad (6)$$

$$\mathcal{P} \xrightarrow{\rho} \oslash, \quad (3) \qquad\qquad \mathcal{P} \xrightarrow{\rho} \oslash, \quad (7)$$

$$\mathcal{R} + \mathcal{P} \xrightarrow{h_1} \mathcal{RP}, \quad (4) \qquad \mathcal{R} + \mathcal{P}_e \xrightarrow{h_2} \mathcal{RP}_e, \quad (8)$$

$$\mathcal{RP} \xrightarrow{f} \mathcal{R} + \mathcal{P}, (5) \qquad \mathcal{RP}_e \xrightarrow{f} \mathcal{R} + \mathcal{P}_e, \quad (9)$$

The Eqs. (2) and (6) are indicating protein synthesis while protein degradation is indicated by Eqs. (3) and (7). The gene switching from ON to OFF state is indicated by Eqs. (4) and (8) while the opposite transition is presented at Eqs. (5) and (9). The system of effective reactions presented here is very simplified in comparison with the complexity of gene regulation and gene expression in mammals. However, such a simplification is highly necessary for establishing a quantitative description based on exactly solvable models.

A stochastic model for the binary gene with the probabilities of finding the gene at the ON (or OFF) state when n gene products are found inside the cell being denoted by α_n (or β_n) have been presented elsewhere [38, 40]. Hence, the state of the system is determined by two random variables (m, n), with $m =$(ON,OFF) and n being a non-negative integer. Those probabilities can be computed for a specific stochastic process determining their evolution which can be done by means of continuous time Markov processes also known as master equations. The master equations are characterized by a combination of the individual transformations that change the state of the system. The left-hand side of a master equation has the rate of change of the probability for the system being in a given state while the right-hand side has the processes that are causing the probabilities to change. At the left-hand side of the master equation, a positive contribution means that the transformation brings the system to current state while transformations taking the system from the current state towards a different one give negative contributions.

The master equations governing the dynamics of the probabilities (α_n, β_n) are written below. We provide an interpretation for the first term on the right-hand side and the remaining can be interpreted in the same framework. The term proportional to k has a positive component α_{n-1} and α_n as a negative component. The former means that if the state of the system is (ON,$n-1$) and there is a synthesis of a gene product the system reaches state (ON,n) while the second means that a synthesis takes the system from current state (ON,n) towards state (ON,$n+1$). The master equation is written as

$$\frac{d\alpha_n}{dt} = k(\alpha_{n-1} - \alpha_n) + \rho[(n+1)\alpha_{n+1} - n\alpha_n]$$
$$- (h_1 n + h_2)\alpha_n + f\beta_n, \qquad (10)$$

$$\frac{d\beta_n}{dt} = \rho[(n+1)\beta_{n+1} - n\beta_n]$$
$$+ (h_1 n + h_2)\alpha_n - f\beta_n, \qquad (11)$$

where the self-repressing gene is modeled considering

$h_1 \neq 0$ and $h_2 = 0$, such that the switching rate from ON to OFF state depends on n. Note that in the self-repressing gene the ON to OFF switching rate has linear dependence on n. The contrary condition, $h_1 = 0$ and $h_2 \neq 0$, results in a model for the externally regulated gene. The solution to the Eqs. 10 and 11 have been obtained exactly for the self-repressing gene [38, 39] and the externally regulated gene [40, 41]. For the two rates h_1 and h_2 being non-null one has a model for a gene that is both self-repressing and externally regulated. That will not be approached here because it is a subject for additional research.

A striking feature of biological organisms is their capability of regulation that ensures that a given gene will be expressed in proper quantities with spatial and temporal precision. Hence, though variation in gene amounts is observed these fluctuations are within specific ranges in normally behaving biological systems. An important question is to find regulatory strategies underpinning such a precision such that a classification of regulatory strategies and their biological function would emerge. For example, there are experiments demonstrating that self-repressing genes are responsible for reducing random fluctuations in gene expression [21, 22, 24]. Indeed, it has been shown that self-repression induces noise reduction such that one may obtain sub-Fano probability distributions [23]. However, the mechanisms enabling such a reduction in noise were not clear. That has appeared later by writing the Fano factor as

$$\mathcal{F} = 1 + \frac{\xi}{\langle n \rangle}, \tag{12}$$

where ξ is the covariance between the variables (m, n) with the values of m (ON,OFF) being represented by corresponding synthesis rates $(k/\rho, 0)$ such that ξ can be computed. The self-repressing gene has the possibility of $\xi < 0$ and sub-Fano probability distributions will occur [1]. Those regimes will happen when the prevailing process of the system is the gene switching between the ON and OFF states while few synthesis or degradation of gene products happen during a given time interval.

The left graph of Fig. 2 shows the Fano factor for a self-repressing gene in the sub-Fano regime. Note the existence of the possibility of finding arbitrarily low values for $\mathcal{F}$ when $\langle n \rangle = 1$. That corresponds to a kinetics with the regulatory protein having a high affinity the regulatory region controling the expression of the gene. At that regime, once a regulatory protein is released from the DNA and the gene becomes ON, an available protein rapidly binds to the DNA and the gene switches back to the OFF state.

The cartoon that we are considering for gene regulation may seem a strong simplification of the the whole picture in metazoans. However we may use that approach for description of eukaryotes under specific assumptions. For example, during its early developmental stages *D. melanogaster* embryos are characterized by a syncytium such that the cells only have their nuclei. That enables us
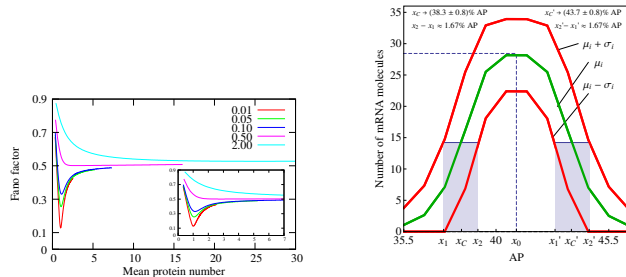

FIG. 2: *Left graph.* Fano factor versus average protein number for the self-repressing gene. The value of a is fixed as 500. The values of b^s are indicated within the graph while we have varied the value of z_0. *Right graph.* Spatial profile of mRNA's average amounts along AP axis of a *D. melanogaster* embryo. We also include the fluctuation on the position of the borders of the peak of expression accordingly with the standard deviation of n at each nucleus along AP axis. The position of the borders are computed at the point where $\langle n \rangle$ is half of its maximal value at position 41.5 % of the embryo length.

to apply the gene transcription model for an externally regulating gene and use it as a first step to understand how spatial noise appears because of noise on amounts of gene products.

Indeed, we carried out that approach to model the *even-skipped (eve)* gene which is important for the formation of spatial protein concentration patterns that in the adult organism will determine specific segments with different functions [2]. *eve* mRNA's spatial pattern is characterized by a Gaussian profile at the onset of gastrulation (see right graph of Fig. 2). To apply our model we assumed a one-dimensional lattice where each node has a single copy of *eve*. The lattice represents the AP axis of the embryo and theoretical values for $\langle n \rangle$ at each node of the lattice were compared with observed values for the fluorescence of mRNA of *eve* as obtained experimentally [42]. At this stage the challenge was to propose a method for converting the detected intensity of immunofluorescence into mRNA numbers. Then we compared the two spatial patterns at the onset of gastrulation (theoretical and experimental) and obtained a good agreement. The second stage was to compute the values of $\langle n \rangle \pm \sigma$ along the whole lattice, where σ indicates the standard deviation on n. Then we compared the position of the borders of the domains and their fluctuations with experimental data [27]. That showed that the theoretical fluctuations were of the same order as of the experimental ones. That is an unexpected result since it indicates that the spatial precision required during mRNA numbers pattern formation in developing embryos can be achieved with a gene transcription regulatory strategy that is not the most precise on controlling transcripts numbers.

III. CELL LEVEL MODELS

There are strong experimental evidence that loss of contact inhibition is a key element underlying the capability of tumors to grow and prevail within a given tissue (forming the *carcinoma in situ*) [33–36] or that hypersensivity to contact inhibition prevents such a prevalence [37]. Those roles indicate the necessity of understanding quantitatively (and geometrically) how contact inhibition affects the dynamics of cell proliferation and the spatial patterns that they form in a given tissue. Such an approach is going to be useful for the understanding of early cancer development and future design of techniques for early cancer diagnosis and treatment. To approach that problem we have recently proposed a co-culture experiment combining keratinocytes (HaCaT or normal) and melanoma (SK-MEL-147 or cancer) cells [3]. We considered an initial configuration of 10:1 (keratinocytes:melanoma cells) and evaluated the cell density daily until confluence was reached. The initial configuration was composed by well-mixed populations of keratinocytes and melanoma cells. At confluence we observed spatial patterns with spreaded normal cells surrounding melanoma cells clusters (see Fig. 3.G). In that experiment the growth of the two sub-populations of cells was tracked and their individual curves are easily adjusted by Gompertz logistic-like curves [43] (see fig. 3.B). Gompertz curves are characterized by a sigmoidal shape such that the population density has a slow growth rate at early phase. That growth rate increases until it reaches a maximal value and starts diminishing while cell population density reaches its maximal value asymptotically. The density growth rate is proportional to a constant which is interpreted as the inverse of the cell division time during the early phase of culture experiments when the growth is still approximately exponential. The fitting has shown that the growth rate of both cell types had same value while the final proportions of the two populations has diminished from 10:1 to $\approx 4 : 1$ (Fig. 3.C shows the temporal evolution of this ratio). That result shows a limitation for using Gompertz-like curves to describe experimental results when distinct cell types interact in the same environment. Particularly, those models have been developed in the context of predator-prey interacting in an environment with finite resources availability [44]. We did not found in literature any instance of co-culture of melanoma and normal cells interacting similarly to a predator-prey system. Hence, a different approach to describe our system quantitatively was required.

To obtain a quantitative description of our model we consider the a variation of the Widom-Rowlinson model [45–47]. A cartoon of our model is presented in Fig. 4.A where the tissue is represented as a two-dimensional grid of size $L \times L$. Melanoma (or normal) cells, indicated in red (or blue), can occupy the grid's vertices with the distance between two cells being the smallest amount of edges connecting their vertices. The latter enable us to define con-
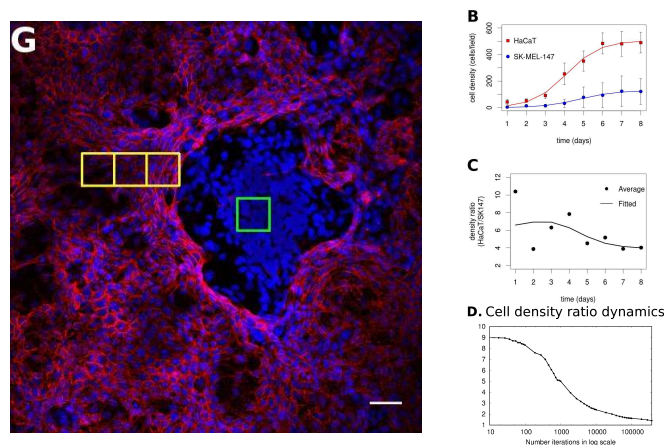

FIG. 3: Graph G. shows a representative configuration of the co-culture experiment at the confluence regime. Graph B. gives the evolution of the individual populations along until confluence is reached and their fitting by a sigmoidal curve. Graph C. the experimental ratio of melanoma to normal cells along time is shown. Graph D. the simulation of the ratio of melanoma to normal cells along time is shown.

tact inhibition by means of an exclusion diameter around the cell as it is represented by the shadowed areas around the red circles of Fig. 4.A. The vertices within the purple regions cannot be occupied by the melanomas while the normal cells cannot occupy the vertices within the red shadowed areas. The exclusion diameters are reciprocal and notice that the melanoma cells occupy their immediate neighbors. In our model the cell type i ($i = 1$ or $i = 2$, respectively, for melanoma or normal) undergo division (with rate α_i), quiescence (with rate σ_i), death (with rate ρ_i) and migration (with rate δ_i) with specific rates.

In Ref. [3] the dynamics of the model is established by means of a Monte Carlo Markov Chain which is briefly described below. A vertex x of the grid is selected with probability L^{-2} and its state is verified. (*1.*) For the vertex being **occupied** by the i-th cell type: the cell is quiescent, *i.e.* x remains occupied, with probability α_i/Q; there is a probability $(n_{\mathrm{e}}^{(i)}(x)\delta_i + \rho_i)/Q$ for vertex x turn empty, where $Q = Q(x) = \alpha_i + n_{\mathrm{e}}^{(i)}\delta_i + \rho_i$. The vertex x turns empty because of cell *death* – with probability ρ_i/Q – or *migration* – with probability $n_{\mathrm{e}}^{(i)}(x)\delta_i/Q$. In case of *migration* the cell arrive at any vertex at distance $D(i,i)$ that satisfies the admissibility rule. The number of vertices around x that can receive the cell is indicated by $n_{\mathrm{e}}^{(i)}(x)$ and $1/n_{\mathrm{e}}^{(i)}(x)$ is the probability for one of these vertices receive the migrating cell. (*2.*) For the vertex x being **empty**: it remains empty with probability $(\rho_1 + \rho_2)/R^1$; the vertex x may become occupied by

———

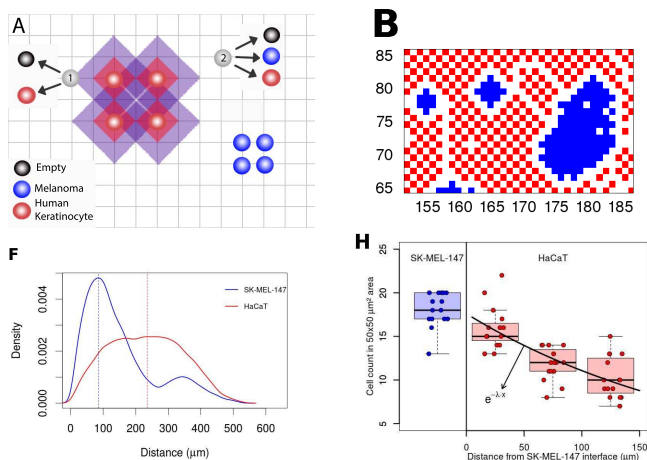[1] Although that probability might be defined by means of a differ-

FIG. 4: Graph A. gives a cartoon of our model for proliferation under different allelophylic degrees. Graph B. is a spatial configuration achieved in our simulations at the co-culture confluence regime. Graph F. gives the experimental cell-to-cell distance distribution (the blue (red) curve indicates cancer (normal) cells histogram). Graph H. shows how normal-to-normal cell distance as they are further away from the interface with the melanoma clusters.

the i-th cell type with probability $n_i(x)(\alpha_i + \delta_i)\mathbb{I}(i)/R$. The occupation of x may happen because of a cell *division* – having probability $n_i(x)\alpha_i\mathbb{I}(i)/R$ – or *migration* – having probability $n_i(x)\delta_i\mathbb{I}(i)/R$, where $R = R(x) = \sum_{i=1}^{2}[\rho_i + n_i(x)(\alpha_i + \delta_i)\mathbb{I}(i)]$. The amount of allowed nearest vertices occupied by the i-th cell type around x is denoted by $n_i(x)$. A migrating cell has a probability $1/n_i(x)$ to be chosen and, after migration, the vertex from which the cell comes from is turned empty. The symbol $\mathbb{I}(i)$ is equal to 1 or 0 when the vertex x, respectively, can or cannot receive the i-th cell type accordingly with admissibility rule. Note that x remains empty with probability one for $R = \rho_1 + \rho_2$.

Fig. 3.D shows the results of our simulations for the ratio between the densities of normal to tumor cells. We show that that ratio follows the same pattern as observed in our experiments. Furthermore, the densities of the two sub-populations also follow the same pattern (as it may be verified in Fig. [3]) such that the kinematics obtained with our model adjusts to that observed experimentally. Hence, it is fair to conclude that our model is a strong candidate to quantify how different degrees of contact inhibittion affects the dynamics of cell proliferation in co-culture experiments, and, maybe, in tissues.

Indeed, fig. 4.B shows a spatial configuration that we obtain after simulating our model. The blue circles form

_____

ent rate our choice avoided introducing more parameters to our simulations.

clusters that are surrounded by the red ones. This pattern is similar to the ones we have observed in our co-culture experiments, fig. 3.G. Furthermore, we also evaluated the typical cell to cell distances observed in our experimental images. Fig. 4.F shows that at confluence regime normal to normal cells typical distance is twice that between cancer cells. That reinforces of our geometrical interpretation of contact inhibition by means of an exclusion diameter and shows the cancer cells as *allelophylic* (*allelo*, the other; *phylia* affinity). In our study we also found the correspondence between the spatial scales of the simulation and the experimental ones. That has enabled us to compare the melanoma clusters characteristics observed experimentally with those of the simulations [3].

IV. PERSPECTIVES

Our results raises perspectives for further research, and a non-exhaustive set of possibilities is discussed below. The use of a stochastic binary model for gene expression on the description of the formation of stripe two of *eve* along the AP axis of the *fruitfly* embryo indicates the necessity of understanding the effects of fluctuations on molecular numbers on the spatial organization of cells. On the other hand, the use of model for a self-repressing gene can be used in cancer context to investigate the behavior of BACH1 production under influence of a biometallic compound such as heme [48]. BACH1 is a transcription factor with negative self-regulation [48] found overexpressed in triple-negative breast cancer cells. Since it has been demonstrated its role as a metastasis promoter, to model how metastic breast cencer cells keep it at higher levels is relevant to the development of new therapuetic approaches to this condition. Heme accelerates BACH1 decay and the model may help on the design of treatement strategies that also reduce both the expected amounts of BACH1 within the cell and their fluctuations such that phonotype heterogeneity of tumor cells is reduced. That reduction has the potential of increasing the cancer treatments effectiveness and may establish guidelines to reduce the invasive and metastatic capabilities in cancer cells. Under a more theoretical perspective, we may also consider the investigation of the meaning of the symmetries of the stochastic binary models [23, 26] aiming to model two interacting genes.

Another possibility of cell levels models is to propose Markov chains to approach tumor heterogeneity at the phenotype level. The tumor progression can yield changes in its architecture that lead tumor cells to die or to develop invasive phenotypes because of scarcity of space and resources [49–52]. Additionally, environmental cues may regulate expression of transcription factors that regulate the internal cell dynamics [53–56] from proliferative towards invasive. Those facts suggest the proposition of a cell level model for tumor progression having two cell phenotypes, with the invasive being originated

from the proliferative, and population dependent transition rates as an effective homeostatic mechanism. On more specific possibilities, the use of the stochastic model for contact inhibition may also be extended to the condition when there are three or more interacting cell types. In that case, we may start with a three cell states, accounting for keratinocytes, melanocytes, and melanomas. Here, the melanomas would result from a modification of melanocytes and we may use those results to investigate the conditions for the progression of a melanoma *in situ* from a normal configuration. The cell level approaches have investigated the two dimensional grids for a description of culture experiments. Our next step is to construct three dimensional grids to enable us to describe *in vivo* experiments and, hence, obtain a richer picture about carcinogenesis. One natural challenge of such an approach is to establish the grid's topology such that different first neighbors rules would come up. Those new simulations would enable us to develop new imaging analysis tools that will be useful for a quantitative characterization of the spatial patterns that are generated at different stages of carcinogenesis. Those characterization may motivate the development of automated tools for tissue characterization by patologists.

Fig. 3.G shows four yellow squares, one within a melanoma cluster and the remaining three sequentially located from the interface of normal and melanoma cells domains. Inspection is sufficient to verify that normal cells nearby melanomas are closer to each other than those at a longer distance from the cluster. That pattern repeats in our experiments and a graph of the densities of keratinocytes at different distances is shown in Fig. 4.H. Box-plots for melanoma-to-melanoma (or normal-to-normal) cell densities at different distances are presented. We show that the cell-to-cell separation falls exponentially which suggests the existence of a molecular mechanisms determining that separation which concentration is dependent of the presence of the melanoma cells. That points out towards the necessity of combining our two approaches for cancer biology at both molecular and cellular levels. That would be useful to determine how molecular level fluctuations give raise to cancer heterogeneity.

Our investigations on molecular mechanisms of carcinogenesis may also have implications for the analysis of low dose and low dose rates rates random effects [57, 58]. It is estimated that radiation therapy accounts for 50% of cancer treatment cases and that application may be playing a role on late appeerance of tumors. Hence, understanding how low dose and low dose rates of ionizing radiation are affecting cells may be an important scientific problem with clinic implications. At those regimes one expects the effects of ionizing radiation to be stochastic such that the it is natural to employ an approach based on probabilistic theory. Initial attempts for stochastic modeling of biological effects are based on target theory [59] while more detailed deterministic models have been proposed recently to account for DNA repair mechanisms of mammal cells [60]. For the latter, we will employ the Langevin technique to evaluate randomness in deterministic models. On a different research direction, one may notice that 90% of radiation treatments use radiofrequency-driven linear accelerators of electrons (RF-Linac). These RF sources are not very precise and may lead radiation to hit both the tumor and neighboring cells. Hence, there are continuous efforts for finding new radiation sources including laser accelerated electrons for the generation of tunable and quasi-energetic x-ray sources [61, 62]. The development of that technology relies hardly on computational simulations of the interaction of laser plasma interactions devoted to the optimization of the generation of x-rays through electron acceleration by lasers. Those simulations may guide experimental designes which would result on the proper generation of x-rays for clinical purposes.

AUTHOR CONTRIBUTIONS

AUS, MFV, MYS, WWL, ASQ, MCCM, collected data, wrote the manuscript and approved the final version of the manuscript to be published. AFR designed the study, collected data, interpreted the results, wrote the manuscript and approved the final version of the manuscript to be published.

[1] Ramos AF, Hornos JEM, Reinitz J. Gene regulation and noise reduction by coupling of stochastic processes. Phys Rev E. 2015;91:020701(R). https://doi.org/10.1103/PhysRevE.91.020701.

[2] Prata GN, Hornos JEM, Ramos AF. Stochastic model for gene transcription on *Drosophila melanogaster* embryos. Phys Rev E. 2016;93:022403. https://doi.org/10.1103/PhysRevE.93.022403.

[3] Morais MC, Stuhl I, Sabino AU, Lautenschlager WW, Queiroga AS, Tortelli Jr TC, et al.. Stochas-

tic model of contact inhibition and proliferation of melanoma *in situ*. Sci Rep. 2017;7:8026. http://doi.org/10.1038/s41598-017-07553-6.

[4] Delbrück M. Statistical fluctuations in autocatalytic reactions. J Chem Phys. 1940;8:120–124. https://doi.org/10.1063/1.1750549.

[5] Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci USA. 2001;98:8614–8619. https://doi.org/10.1073/pnas.151588598.

[6] Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297:1183–1186. https://doi.org/10.1126/science.1070919.

[7] Simpson ML, Cox CD, Sayler GS. Frequency domain analysis of noise in autoregulated gene circuits. Proc Natl Acad Sci USA. 2003;100:4551–4556. https://doi.org/10.1073/pnas.0736140100.

[8] Blake WJ, Kaern M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. Nature. 2003;422:633–637. https://doi.org/10.1038/nature01546.

[9] Raser JM, O'Shea EK. Noise in gene expression: origins, consequences, and control. Science. 2005;309:2010–2013. https://doi.org/10.1126/science.1105891

[10] Arias AM, Hayward P. Filtering transcriptional noise during development: concepts and mechanisms. Nat Rev Genet. 2006;7:34–44. https://doi.org/10.1038/nrg1750.

[11] Hooshangi S, Weiss R. The effect of negative feedback on noise propagation in transcriptional gene networks. Chaos. 2006;16:026108. https://doi.org/10.1063/1.2208927.

[12] Chubb JR, Trcek T, Shenoy SM, Singer RH. Transcriptional Pulsing of a Developmental Gene. Curr Biol. 2006;16:1018–1025. https://doi.org/10.1016/j.cub.2006.03.092.

[13] Boettiger AN, Levine M. Synchronous and Stochastic Patterns of Gene Activation in the *Drosophila* Embryo. Science. 2009;325:471–3. https://doi.org/10.1126/science.1173976

[14] Paré A, Lemons D, Kosman D, Beaver W, Freund Y, McGinnis W. Visualization of individual Scr mRNAs during Drosophila embryogenesis yields evidence for transcriptional bursting. Curr Biol. 2009;19(23):2037–2042. https://doi.org/10.1016/j.cub.2009.10.028

[15] Crudu A, Debussche A, Radulescu O. Hybrid stochastic simplifications for multiscale gene networks. BMC Syst Biol. 2009;3(89). https://doi.org/10.1186/1752-0509-3-89

[16] Munsky B, Neuert G, van Oudenaarden A. Using Gene Expression Noise to Understand Gene Regulation. Science. 2012;336(6078):183–187. https://doi.org/10.1126/science.1216379.

[17] Chalancon, G, Ravarani, C, Balaji, S, Martinez-Arias, A, Aravind, L, Jothi, R, et al.. Interplay between gene expression noise and regulatory network architecture. Trends Genet. 2012;28:221–232. https://doi.org/10.1016/j.tig.2012.01.006

[18] Sanchez A, Golding I. Genetic Determinants and Cellular Constraints in Noisy Gene Expression. Science. 2013;342(6163):1188–1193. https://doi.org/10.1126/science.1242975.

[19] Little SC, Tikhonov M, Gregor T. Precise Developmental Gene Expression Arises from Globally Stochastic Transcriptional Activity. Cell. 2013;154(4):789–800.

[20] Jiang P, Ludwig MZ, Kreitman M, Reinitz J. Natural variation of the expression pattern of the segmentation gene even-skipped in melanogaster. Dev Biol. 2015;405(1):173–181. https://doi.org/10.1016/j.ydbio.2015.06.019

[21] Savageau MA. Comparison of classical and autogenous systems of regulations in inducible operons. Nature. 1974;252:546–549. http://dx.doi.org/10.1038/252546a0

[22] Becskei A, Serrano L. Engineering stability in gene networks by autoregulation. Nature. 2000;405(6786):590–3. https://doi.org/10.1038/35014651.

[23] Ramos AF, Hornos JEM. Symmetry and stochastic gene regulation. Phys Rev Lett. 2007;99(10):108103. https://doi.org/10.1103/PhysRevLett.99.108103.

[24] Nevozhay D, Adams RM, Murphy KF, Josic K, Balazsi G. Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. Proc Natl Acad Sci USA. 2009 Mar;106(13):5123–5128. http://dx.doi.org/10.1073/pnas.0809901106.

[25] Shahrezaei V, Swain PS. Analytical distributions for stochastic gene expression. Proc Natl Acad Sci USA. 2008;105:17256–17261. http://doi.org/10.1073/pnas.0803850105.

[26] Ramos AF, Innocentini GCP, Forger FM, Hornos JEM. Symmetry in biology: from genetic code to stochastic gene regulation. IET Syst Biol. 2010;4:311–329. http://doi.org/10.1049/iet-syb.2010.0058.

[27] Surkova S, Kosman D, Kozlov K, Manu, Myasnikova E, Samsonova A, et al.. Characterization of the *Drosophila* Segment Determination Morphome. Dev Biol. 2008;313(2):844–862. https://doi.org/10.1016/j.ydbio.2007.10.037.

[28] Barr KA, Martinez C, Moran JR, Kim AR, Ramos AF, Reinitz J. Synthetic enhancer design by in silico compensatory evolution reveals flexibility and constraint in cis-regulation. BMC Syst Biol. 2017;11:116. http://doi.org/10.1186/s12918-017-0485-2.

[29] Fraser D, Kaern M. A chance at survival: gene expression noise and phenotypic diversification strategies. Mol Microbiol. 2009;71:1333–1340. http://doi.org/10.1111/j.1365-2958.2009.06605.x.

[30] Balazsi G, van Oudenaarden A, Collins JJ. Cellular decision making and biological noise: from microbes to mammals. Cell. 2011;144:910–925. http://doi.org/10.1016/j.

[31] Ben-Jacob E, Coffey DS, Levine H. Bacterial survival strategies suggest rethinking cancer cooperativity. Trends Microbiol. 2012;20:403–410. http://doi.org/10.1016/j.tim.2012.06.001.

[32] Mitosch K, Rieckh G, Bollenbach T. Noisy Response to Antibiotic Stress Predicts Subsequent Single-Cell Survival in an Acidic Environment. Cell Systems. 2017;4:393–403. http://doi.org/10.1016/j.cels.2017.03.001.

[33] Abercrombie M. Contact inhibition and malignancy. Nature. 1979;281(5729):259–262. https://www.ncbi.nlm.nih.gov/pubmed/551275

[34] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–674. https://doi.org/10.1016/j.cell.2011.02.013

[35] Stockinger A, Eger A, Wolf J, Beug H, Foisner R. E-cadherin regulates cell growth by modu-

https://doi.org/10.1016/j.cell.2013.07.025

lating proliferation-dependent beta-catenin transcriptional activity. J Cell Biol. 2001;154(6):1185–1196. https://doi.org/10.1083/jcb.200104036

[36] Puliafito A, Hufnagel L, Neveu P, Streichan S, Sigal A, Fygenson DK, et al.. Collective and single cell behavior in epithelial contact inhibition. Proc Natl Acad Sci USA. 2012;109(3):739–744. https://doi.org/10.1073/pnas.1007809109

[37] Seluanov A, Hine C, Azpurua J, Feigenson M, Bozzella M, Mao Z, et al.. Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. Proc Natl Acad Sci USA. 2009;106(46):19352–19357. https://doi.org/10.1073/pnas.0905252106

[38] Hornos JE, Schultz D, Innocentini GC, Wang J, Walczak AM, Onuchic JN, et al.. Self-regulating gene: an exact solution. Phys Rev E. 2005;72(5):051907. https://doi.org/10.1103/PhysRevE.72.051907

[39] Ramos AF, Innocentini GCP, Hornos JME. Exact time-dependent solutions for a self-regulating gene. Phys Rev E. 2011;83:062902. https://doi.org/10.1103/PhysRevE.83.062902

[40] Innocentini GC, Hornos JE. Modeling stochastic gene expression under repression. J Math Biol. 2007;55(3):413–31. https://doi.org/10.1007/s00285-007-0090-x

[41] Iyer-Biswas S, Hayot F, Jayaprakash C. Stochasticity of gene products from transcriptional pulsing. Phys Rev E. 2009;79:031911. https://doi.org/10.1103/PhysRevE.79.031911

[42] Janssens H, Hou S, Jaeger J, Kim A, Myasnikova E, Sharp D, et al.. Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster even-skipped* gene. Nat Genet. 2006;38:1159–1165. https://doi.org/10.1038/ng1886

[43] Benzekry S, Lamont C, Beheshti A, Tracz A, Ebos JML, Hlatky L, et al.. Classical mathematical models for description and prediction of experimental tumor growth. PLoS Comput Biol. 2014;10(8):e1003800. https://doi.org/10.1371/journal.pcbi.1003800

[44] Gerlee P. The model muddle: in search of tumor growth laws. Cancer Res. 2013;73(8):2407–2411. https://doi.org/10.1158/0008-5472.CAN-12-4355

[45] Widom B, Rowlinson JS. New model for the study of liquid-vapor phase transitions. J Chem Phys. 1970;52(4):1670–1684. https://doi.org/10.1063/1.1673203

[46] Mazel A, Y Suhov Y, Stuhl I, Zohren S. Dominance of most tolerant species in multi-type lattice Widom-Rowlinson models. J Stat Mech. 2014;2014:P08010. http://stacks.iop.org/1742-5468/2014/i=8/a=P08010

[47] Mazel A, Suhov Y, Stuhl I. A Classical WR Model with q Particle Types. J Stat Phys. 2015;159(5):1040–1086. https://doi.org/10.1007/s10955-015-1219-8

[48] Lee J, Lee J, Farquhar KS, Yun J, Frankenberger CA, Bevilacqua E, et al.. Network of mutually repressive metastasis regulators can promote cell heterogeneity and metastatic transitions. Proc Natl Acad Sci USA. 2014;111(3):E364–E373. https://doi.org/10.1073/pnas.1304840111.

[49] Gatenby RA, Gillies RJ. Why do cancers have high aerobic glycolysis? Nat Rev Cancer. 2004 Nov;4(11):891–899. https://doi.org/10.1038/nrc1478.

[50] Smallbone K, Gatenby RA, Gillies RJ, Maini PK, Gavaghan DJ. Metabolic changes during carcinogenesis: potential impact on inva-

siveness. J Theor Biol. 2007;244(4):703–713. https://doi.org/10.1016/j.jtbi.2006.09.010.

[51] Ibrahim-Hashim A, Robertson-Tessi M, Enriquez-Navas PM, Damaghi M, Balagurunathan Y, Wojtkowiak JW, et al.. Defining cancer subpopulations by adaptive strategies rather than molecular properties provides novel insights into intratumoral evolution. Cancer Res. 2017;77(9):2242–2254. https://doi.org/10.1158/0008-5472.CAN-16-2844.

[52] Alfarouk KO, Ibrahim ME, Gatenby RA, Brown JS. Riparian ecosystems in human cancers. Evol Appl. 2013;6(1):46–53. https://doi.org/10.1111/eva.12015.

[53] Hoek KS, Eichhoff OM, Schlegel NC, Døbbeling U, Kobert N, Schaerer L, et al.. In vivo switching of human melanoma cells between proliferative and invasive states. Cancer Res. 2008;68(3):650–656. https://doi.org/10.1158/0008-5472.CAN-07-2491.

[54] Carreira S, Goodall J, Denat L, Rodriguez M, Nuciforo P, Hoek KS, et al.. Mitf regulation of Dia1 controls melanoma proliferation and invasiveness. Genes Dev. 2006;20(24):3426–3439. https://doi.org/10.1101/gad.406406.

[55] Goodall J, Carreira S, Denat L, Kobi D, Davidson I, Nuciforo P, et al.. Brn-2 represses microphthalmia-associated transcription factor expression and marks a distinct subpopulation of microphthalmia-associated transcription factor–negative melanoma cells. Cancer Res. 2008;68(19):7788–7794. https://doi.org/10.1158/0008-5472.CAN-08-1053.

[56] Haass NK, Beaumont KA, Hill DS, Anfosso A, Mrass P, Munoz MA, et al.. Real-time cell cycle imaging during melanoma growth, invasion, and drug response. Pigment Cell Melanoma Res. 2014;27(5):764–776. https://doi.org/10.1111/pcmr.12274.

[57] Brenner DJ. The use of the linear-quadratic model in clinical radiation oncology can be defended on the basis of empirical evidence and theoretical argument. Med phys. 1997;24:1245–8. https://doi.org/10.1118/1.598086.

[58] Bodgi L, Canet A. L'énigme de l'interprétation biologique du modéle linéaire-quadratique enfin résolue? Une synthése pour les non-mathématiciens. Cancer Radiother. 2016;20:314–321. https://doi.org/10.1016/j.canrad.2016.02.014.

[59] Bodgi L, Canet A. Mathematical models of radiation action on living cells: From the target theory to the modern approaches. A historical and critical revew. J Theor Biol. 2016;394:93–101. .

[60] Wodarz D, Sorace R. Dynamics of cellular responses to radiation. PLoS Comput Biol. 2014;10(4):e1003513. https://doi.org/10.1371/journal.pcbi.1003513.

[61] Powers, ND, Ghebregziabher, I, Golovin, F, Liu, C, Chen, S, Banerjee, S, et al.. Quasi-monoenergetic and tunable X-rays from a laser-driven Compton light source. Nat Photonics. 2014;8:28–31. https://doi.org/10.1038/nphoton.2013.314.

[62] Giulietti, A Editor. Laser-Driven Particle Acceleration Towards Radiobiology and Medicine. New York, NY: Springer Berlin Heidelberg; 2016. https://doi.org/10.1007/978-3-319-31563-8.

FIG. 1: The left cartoon is a representation of a self-repressing gene while the right cartoon represents an externally regulated gene. (repeated from page 3)

FIG. 2: *Left graph.* Fano factor versus average protein number for the self-repressing gene. The value of a is fixed as 500. The values of b^s are indicated within the graph while we have varied the value of z_0. *Right graph.* Spatial profile of mRNA's average amounts along AP axis of a *D. melanogaster* embryo. We also include the fluctuation on the position of the borders of the peak of expression accordingly with the standard deviation of n at each nucleus along AP axis. The position of the borders are computed at the point where $\langle n \rangle$ is half of its maximal value at position 41.5 % of the embryo length. (repeated from page 4)

FIG. 3: Graph G. shows a representative configuration of the co-culture experiment at the confluence regime. Graph B. gives the evolution of the individual populations along until confluence is reached and their fitting by a sigmoidal curve. Graph C. the experimental ratio of melanoma to normal cells along time is shown. Graph D. the simulation of the ratio of melanoma to normal cells along time is shown. (repeated from page 5)

FIG. 4: Graph A. gives a cartoon of our model for proliferation under different allelophylic degrees. Graph B. is a spatial configuration achieved in our simulations at the co-culture confluence regime. Graph F. gives the experimental cell-to-cell distance distribution (the blue (red) curve indicates cancer (normal) cells histogram). Graph H. shows how normal-to-normal cell distance as they are further away from the interface with the melanoma clusters. (repeated from page 6)