W. Allaerts [2014] Med. Hypotheses Res. 9: 1-15.

REVIEW & HYPOTHESIS

Annotation Grids for Cell Lineage and Tumor Metastasis: II. Defining the Architecture for Multiscale Probabilistic and Deterministic Models

Wilfried Allaerts*

Biological Publishing, P. O. Box 104, NL-7440 AC Nijverdal, The Netherlands

Abstract. The versatility of cell groups, either by malignant transformation or controlled reprogramming, forms an important challenge to modeling studies for cellular development. Especially in tumor metastasis, multiscale modeling benefits from the combination of probabilistic and deterministic aspects of cellular behavior. The theoretical notions used in multiscale modeling are summarized, starting from a cell-free finite state projection model to which methods for cell-based cell modeling are added. From recent studies addressing the notions of cell developmental 'roadmaps' and cell-state 'landscapes', a comprehensive architecture is proposed for the study of cell-state modeling.

Correspondence: Dr. Wilfried Allaerts, Biological Publishing, P. O. Box 104, NL-7440 AC Nijverdal, The Netherlands. E-Mail: w.allaerts@planet.nl

Received on August 20, 2013; accepted on April 24, 2014.

1. Introduction

The versatility of cell groups, characterized by their histological features, or defined by their developmental origin and fate, or demonstrated by their clinical or pathogenic behaviour, forms a persistent problem in cell biology and related disciplines (Allaerts, 2008, 2011). Algorithms for cell class discovery and class prediction, based on high-throughput gene expression monitoring, have been introduced in the pathology of leukemia by Golub and group (Golub *et al.*, 1999). Malignancy of neoplasm formation recently has been comprehensively correlated to the metastatic behaviour of cells derived from cancer stem cells. According to Nguyen (2011), neoplastic development into metastasis may be explained in terms of an escape from homeostatic control following a cellular adaptive response to a stressful environment. Cells escaping their environment do so by evading into the vascular system and migration into characteristic distant organs (Hess *et al.*, 2006). The behaviour of metastasis-propagating cells follows a stochastic process with a relatively rare prevalence: according to Nguyen (2011) this is exemplified by the low ratio of metastases formed relative to the vast amount of tumor cells shed daily into the circulation.

On the other hand, the reprogramming of differentiated somatic cells into a pluripotent state was shown to be inducible by transduction of only four defined transcription factors (Takahashi *et al.*, 2007). Moreover, the induced pluripotent stem cells (iPS) appear to be able to differentiate into cell types of the three germ layers in vitro (ectoderm, mesoderm and endoderm) and into teratomas (Takahashi *et al.*, 2007). Reprogramming of somatic cells into iPS appears to occur in waves (Polo *et al.*, 2012). These authors described the establishment of bivalent domains after a first transcriptional wave and identify genes that act as 'roadblocks' in reprogramming. Changes in DNA methylation take place after a second reprogramming wave when cells acquire stable pluripotency (Polo *et al.*, 2012). A more complex 'seesaw' model is proposed by Shu *et al.* (2013). Hereby, a balance is established between mesendodermal (ME) and ectodermal (ECT) lineage specification or pluripotency, where reprogramming is facilitated by pluripotency factors and/or counteracting lineage specifiers (Shu *et al.*, 2013).

Also for the understanding of the differentiation of cancer stem cells into metastatic cells, it appears that a minimal number of transcription factors and necessary environmental cues are essential for allowing metastatic behaviour. Nevertheless, metastatic behaviour is still poorly understood and a capital burden to worldwide health care.
The study of integrative cancer systems shows that different biological scales are important to understand tumor development and metastasis, from initial molecular mechanisms at the transcription level until the clinically observed dysfunctioning of vital organs at the final stage. Therefore a multiscale cancer model is needed to quantify the processes during tumour development (Wang *et al.*, 2011). At the level of biochemical reaction systems, derivative-based sensitivity analysis has been shown to be rather inappropriate in order to consider the influence of multiple biochemical factors on the system response (Zhang *et al.*, 2009). Zhang and co-workers follow a probabilistic model, using a variance-based sensitivity analysis technique that takes into account basic biophysical and thermodynamic properties underlying a biochemical reaction system. However, biological cell systems can hardly be considered as homogeneous mixtures of molecular species, but show different levels of compartmentalization and regulatory

processes at different time scales. Wang and co-workers (2007) developed a multiscale model of non-small cell lung cancer using 2D virtual micro-environments in order to simulate spatio-temporal expansion of tumors. Moreover, cellular phenotype decision algorithms and Michaelis-Menten kinetics are used to simulate cell proliferation, migration, quiescence and cell death phenomena (Wang *et al.*, 2007).

The theoretical background and definitions of mutiscale modeling are given in paragraph 2. First, the outlines of a cell-free fully stochastic modeling are summarized. The notions of 'propensity matrix' and general master equation are adapted from the Finite State Projection (FSP) algorithm, developed by Munsky and Khammash (2006). Then, algorithms for cellular and higher scale modeling are explained, like 'random walk' in a 2D-lattice and 'search precision' algorithms (after Mansury and Deisboeck, 2003), incorporating both stochastic and deterministic elements. Finally, statistics for cellular heterogeneity and single-cell expression analysis are discussed, in order to define roadmap marks and guidelines for studying cell differentiation and/or reprogramming. Correspondence analysis and shape analysis of gene expression patterns are modeled according to the reprogramming studies of Hochedlinger and group (Polo *et al.*, 2012).

Homeostatic cell control is based on a few important principles of living organization, namely the compartmentalization, stable phenotypic expression and communication between the different compartments and their environment (Yamasaki and Nauss, 1996; Mercier and Hatton, 2001). On the other hand, the role of morphogenetic gradients in normal development (*e.g.*, Dasen *et al.*, 1999) and the characterization of core pathways of neoplasm formation (The Cancer Genome Atlas Research Network, 2008) have highlighted the versatility of the cellular differentiation system. The well-known genotype-phenotype trade-off, though at variance within neoplastic cell populations evading homeostatic balance (Nguyen, 2011), shows the importance of 'roadmaps' followed in versatile but stabilized intermediates in cell reprogramming (Polo *et al.*, 2012). In paragraph 3 (SECTION **3.2**) the architectures of a reprogramming and/or neoplastic/metastatic gene expression landscape are proposed.

2. Theoretical Background and Definitions of Multiscale Modeling

2.1. Probabilistic and Deterministic Modeling

Robustness and performance of dynamical systems like biological cells and cell networks recently have been a subject of quantitative and predictive modeling studies from an engineering viewpoint (El-Samad *et al.*, 2005; El-Samad and Khammash, 2006). The design of these biological systems, and even deterministic stability and performance properties of the biological system, may be generated following probabilistic or stochastic models (El-Samad *et al.*, 2006). Probabilistic or stochastic modeling refers to the construction of a dynamic system of complex nonlinear interactions, in which system different metabolism architectures or feedback elements can be designed (translation promotion, feedback and feedforward loops, molecular chaperoning, time delays, redundancy and crosstalk) (El-Samad *et al.*, 2005, 2006). Deterministic modeling (also hierarchical modeling) on the other hand refers to specific reproducible sequences of events in the system that are classically designated as cause-and-effect relationships.

Mostly these events are recognised by observing the dynamical nonlinear interactions at larger distance. Another definition is that 'deterministic events are those events for which the event time is fully determined by the state of the system' (Tindemans, 2009), whereas stochastic events are the result of a random process.

2.2. Cell-Free Stochastic Modeling: Finite State Projection (FSP) Algorithm

In a cell-free approach, Munsky and Khammash (2006) take benefit of the following 'Finite State Projection (FSP) Algorithm'. Consider a volume system of N distinct reacting chemical species. Define $p(x; t)$ the probability that the chemical system will have a molecular population vector x at time t, where $x \in N^N$ is a vector of integers representing a specific population of the N molecular species. So far, no presumptions are made of the homogeneity or inhomogeneity of the population of either of the N molecular species, nor is any compartmentalization imposed on these populations.
Let $a_\mu(x)\,dt$ the non-negative propensity function and v_μ the stoechiometric transition vector for each of the $\mu\ \{1,2,\ldots,M\}$ chemical reactions. Given that we know the probability density vector at t, then the probability that the system will be in the state x at time $t + dt$ is

$$p(x;t+dt) = p(x;t)(1-\sum_{\mu=1}^{M} a_\mu(x)dt) + \sum_{\mu=1}^{M} p(x-v_\mu)a_\mu(x-v_\mu)dt \tag{1}$$

Herein, the first term on the right is the probability that the system begins in state x and will remain there until $t + dt$. The second term reflects the probability that the system transits into x from a different state in the considered time step dt.

From eq. ***(1)*** the following differential equation is derived, known as the 'Chemical Master Equation' (CME) (Munsky and Khammash, 2006):

$$p'(x;t) = \frac{p(x;t+dt)-p(x;t)}{dt} = -p(x;t)\sum_{\mu=1}^{M} a_\mu(x) + \sum_{\mu=1}^{M} p(x-v_\mu;t)a_\mu(x-v_\mu) \tag{2}$$

Combining all reactions that begin or end with state x, the time derivative of the probability density of state x can be written in vector form as:

$$p'(x;t) = \left[-\sum_{\mu=1}^{M} a_\mu(x)a_1(x-v_1)a_2(x-v_2)\ldots a_M(x-v_M) \right] \begin{bmatrix} p(x;t) \\ p((x-v_1);t) \\ p((x-v_2);t) \\ p((x-v_M);t) \end{bmatrix} \tag{3}$$

Now, Munsky and Khammash will a priori fix a sequence of elements in N^N $(x_1,x_2, \ldots)$ and define: $X: = [x_1, x_2, \ldots]^T$.
The particular sequence $x_1, x_2, \ldots$ is chosen to visit every element of the entire space N^N; so, the choice of X corresponds to a particular enumeration of space N^N:

$$P'(X;t) = A \cdot P(X;t) \tag{4}$$

which is an ODE (ordinary differential equation) where *A* is the state reaction matrix and P the complete probability density state vector at time *t*. The state reaction matrix contains information regarding every reaction, each weighted by the corresponding propensity function. The elements of *A* are given as:

$$A = \begin{Bmatrix} -\sum_{\mu=1}^{M} a_\mu(x_i) & for & i = j \\ a_\mu(x_i) & \forall j: & x_j = x_i + x_\mu \\ 0 & otherwise & \end{Bmatrix} \tag{5}$$

A is independent of *t*; all of its diagonal elements are non-positive; all its off-diagonal elements are non-negative; and all its columns sum to exactly zero. The solution to the linear ODE beginning at $t = 0$ and ending at $t = t_f$ is:

$$P(X;t_f) = \Phi(0;t_f) \cdot P(X;0) \tag{6}$$

In the case where there are only a finite number of 'reachable' states, the operator $\Phi\ (0;\ t_f\ )$ is the exponential of $A\ _{tf}$ and one may theoretically compute the solution from:

$$P(X;t_f) = \exp(A_{tf}) \cdot P(X;0) \tag{7}$$

Eq. *(7)* means that theoretically, according to Munsky and Khammash (2006), the 'state' of the system (of *N* molecular species) can be modelled following exponential equations. In the following paragraphs, a chemical interpretation of these general mathematical equations is derived starting from a well-known category of neoplastic transformations, namely the group of non-small cell lung cancer adenomas (Wang *et al.*, 2007).

2.3. Multiscale Definitions and Deterministic Performance Analysis

In order to understand the multiscale character of modeling cancer progression (neoplasm formation and metastasis), a few definitions are indispensable. In the FSP algorithm (see SECTION **2.2**) no compartmentalization of the system is used. The cell, being either of resident type or of a migratory type, forms the first level of compartmentalization. Higher relevant levels of compartmentalization include the histotypical (epithelial, supportive tissue, *etc.*), organ typical and 'segmental' levels (e.g. route of blood and lymphatic drainage) (Allaerts, 2008).

In the model of Wang et al. (2007) the microenvironment is modeled in silico in a two-dimensional lattice. Each cell of the lattice encompasses a self-maintained molecular interaction network, that is constructed to represent cellular progression in non-small cell lung cancer (NSCLC).

Four cellular 'decision phenotypes' are considered in the model of Wang *et al.*

(2007): namely proliferation, migration, quiescence and cell death. Cells may decide to migrate or proliferate, depending on the chemical gradients in their environments. In their model, migration of (cancer) cells depends on surpassing a threshold of 'rate of change' of phospholipase *C*-gamma (PLCγ) in response to epidermal growth factor (EGF) and EGF receptor (EGFR) signaling (Mouneimne *et al.*, 2004). On the other hand, the rate of change of 'extracellular signal-regulated kinase' (ERK) decides if a cell proceeds with proliferation (Brognard and Dennis, 2002; Vicent *et al.*, 2004).

The ODEs for the interaction between different molecules in the cells are of the form:

$$\frac{d(X_i)}{dt} = \sum v_{production} - \sum v_{consumption} \qquad \text{(Wang } et\ al.\text{, 2007)} \qquad (8)$$

where the X_i represent the 20 components of the molecular pathway (starting from EGF signaling, phosphorylation of EGFR and leading to phosphorylation of ERK and activation of the cell cycle). Each biochemical reaction is characterized by a kinetic parameter v_i with forward and reverse rate constants.

Wang *et al.* (2007) take benefit of the 'search precision' algorithm developed by Mansury and Deisboeck (2003). This 'search precision' refers to a statistical average representing "the extent to which tumor cells correctly evaluate the permissibility of a location" (Mansury and Deisboeck, 2003, p. 326). Technically, the search precision parameter Ψ is a parameter between zero and one. When $\Psi = 1$, then it represents the sesarch with 100 % precision, meaning that the tumor cells always correctly evaluate the permissibility of a location without error. When $\Psi = 0$, then a random walk motion is followed at all migration moments. A very important result of Mansury and Deisboeck (2003), is that a fully biased search procedure (with $\Psi = 1$) never in their study emerges as the optimal search strategy. The spatio-temporal expansion of a tumor therefore is faster when the search precision is less than 100%. The maximal biase is towards locations with optimal abundance of nutrients (glucose and O_2), lowest levels of toxicity and least resistance due to degradation of extracellular matrix components. Moreover, their data indicate a prominent phase transition mechanism suggesting that tumor cells do not employ a 100% search precision to attain maximal spatial velocity. Also, "in order to accelerate invasiveness even further, a reduction of the tumor cells proliferation rate is necessary, leading to an even more biased search process" (Mansury and Deisboeck, 2003, p. 337).

2.4. From Annotation to System Roadmaps

Using high throughput technology, and especially using PCR array profiles, an increasing number of cancer types has been analyzed by comparing cancer and normal tissues of large numbers of patients. Consequently, annotation maps for growing numbers of key genes involved in cancer development have been compiled. For instance in case of lung cancer, while the exact (molecular) mechanisms are still under investigation, a list of 84 key genes has been identified (Human Lung Cancer PCR Array, SABiosciences, RT2 Profiler PCR Arrays, Version 4.0, May 25, 2012). In these gene pools, an important

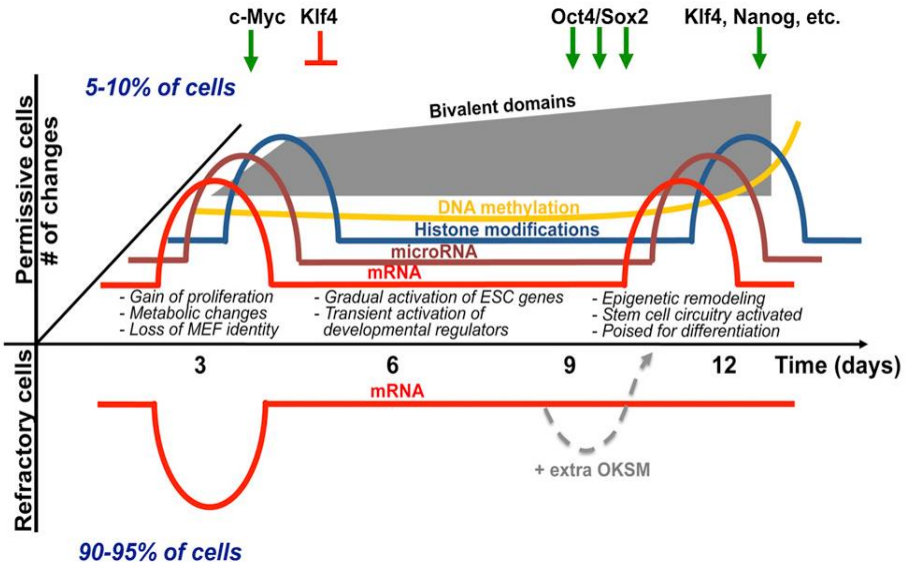

Figure 1. Comprehensive 'roadmap' model presented by Polo et al. (2012). The authors demonstrated that 5-10 % cells are permissive to change and show a biphasic pattern of mRNA/miRNA expression and individual histone marks. They also found that in the permissive cells, bivalent domains were generated after an initial burst, whereas DNA methylation changes occur predominantly at the end of reprogramming. According to this model, c-Myc/Klf4 mostly drive the first phase, whereas Oct4/Sox2/Klf4 drive the second phase. (© 2012, Elsevier Inc., Cell, Vol. 151, p. 1630)

number of genes are involved in tumor suppressor gene inhibition and oncogene activation. These affected tumor suppressors and oncogenes are involved in regulating immune response, apoptosis, cell cycle and cell adhesion pathways. Some of these genes are involved in various pathways, like the above discussed EGFR (P13/AKT signaling, apoptosis, cell cycle regulation). Other genes are associated with the tumor's metastatic potential, like the STAT1 and STAT2 genes (also involved in immune responses and apoptosis), the silencing of cell-cell signaling and first steps of apoptosis induction (annexin V or ANXA5) (Koopman *et al.*, 1994). However, the multiple functional roles of these gene transcripts and the overall redundancy in these transcriptomes urge for further characterization of the interaction patterns between these pathways. Also further modelling is required for clarification of the system roadmaps describing either normal or neoplastic (cancer) development.

In order to study phenotype versatility or reprogramming, it is important to have homogeneous cell populations at one's disposal (see also Allaerts, 2008). Polo *et al.* (2012) used different strategies to determine the degree of cell heterogeneity, such as correspondence analysis of single-cell expression data (Fluidigm technology) and temporal analysis of bimodal or unimodal gene expression patterns. Several technological platforms became available during recent years enabling the analysis of single-cell DNA expression, histone modification and DNA methylation analysis (*e.g.*, EpiTyping), mRNA and miRNA expression, etcetera. Using a combined approach of these technologies, Polo *et al.* (2012) successfully determined reprogramming intermediates poised to becoming iPS cells (**FIG. 1**). The statistical technique used in clustering these intermedi-

ates was based on principal component analysis (see next section).

2.5. Multivariate Statistical Techniques

Multivariate statistical techniques were shown to be powerful methods for analyzing large datasets in geography and other disciplines: in order to identify groups of inter-correlated variables or in order to reduce the number of variables being studied, or simply to rewrite the data set in an alternative form (Johnston, 1980). Two related techniques have become very popular, namely principal component analysis (PCA) and factor analysis, both being a form of causal modeling referred to as 'path analysis' in sociology (Duncan, 1975). Technically the techniques take benefit of the 'eigenvalue problem' in matrix calculus (Gould, 1967; Brinkmann and Klotz, 1971) and 'varimax criterion' for analytic rotation of data sets (Kaiser, 1958). Polo *et al.* (2012) use PCA to unravel the molecular connectivity of genes, e.g. when a cell is progressing from one stage to another, so enabling the identification of reprogramming intermediates.

Briefly, PCA results in rewriting a data matrix, consisting of n variables and N observations, into another *n x n* form of a new set of variables (principal components). Herewith, the new variables have to fulfill two criteria: a) they are weighted representations of the original data set, and b) they are uncorrelated one with another.
Provided the data set P of variables X_i can be transformed into a *n* variable system:

$$X_1 = f(C_I, C_{II}, C_{III}, C_{IV}, \ldots C_n) \quad (9)$$
$$X_2 = f(C_I, C_{II}, C_{III}, C_{IV}, \ldots C_n)$$
$$(\ldots)$$
$$X_n = f(C_I, C_{II}, C_{III}, C_{IV}, \ldots C_n)$$

Then comparison of the importance of each component C_I, C_{II}, … to each variable will indicate the extent of any 'common pattern' among the variables. In general the first (and second) principal component(s) will account for the main part of the variance of the data set. The component loadings of each new variable are based on the calculation of the eigenvalues of the correlation matrix. The eigenvalue (λ) corresponds with the squared loadings for variable *j* on component *i*, or, in other words it accounts for the total variance accounted for by the component *i*, according to:

$$\lambda_i = \sum_{j=1}^{n} L_{ij}^2 \quad (10)$$

where L_{ij} is the loading for variable *j* on component *i* and λ_i is the eigenvalue for component *i* (Johnston, 1980). Geometrically, the correlations between variables and principal components can be derived from the normalized (standardized and rescaled) data set. Given an elliptical cloud of data points in the n-dimensional space, the cosine of the angle between the lines representing two variables equals the correlation between these variables. Also, the largest diameter (main axis) of the cloud and the secondary axis perpendicular to the main axis represent the first and second principal components of the data set.

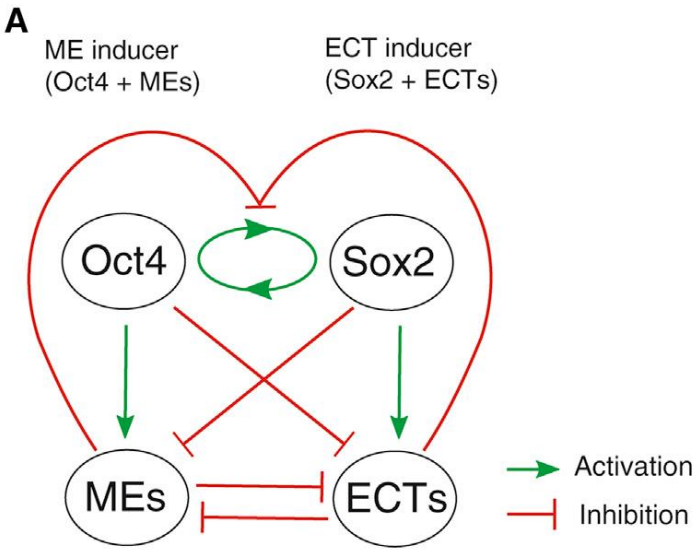

Figure 2. Balanced equilibrium, so-called 'seesaw model', presented by Shu et al. (2013). The seesaw model is used to explain the facilitating of induced pluripotency. The 'pluripotency module' (self-activation of Oct4 and Sox2) is coupled to a 'differentiation module' (mutual antagonism between the ME and ECT lineages). ME: mesendodermal cell lineage; ECT: ectodermal cell lineage. (© 2013, Elsevier Inc., Cell, Vol. 153, p. 972)

Factor analysis differs from the PCA method as follows: each variable is written as a function of n factors and an error term (ε_i) put outside its equation, representing the unique variance for each variable. Practically this means that the principal components have to be normalized. However, if we want to study the temporal variability of principal components (*e.g.*, during development or reprogramming), normalization seems unwanted.

Interestingly, the PCA technique may provide genuine insight into the connectivity of gene transcription products in a given cell type (provided a homogeneous cell population) at a certain stage of development or during reprogramming. However, understanding the implications of a 'reprogramming landscape' (Shu *et al.* 2013) depends on the interactions between the different cell states, as well as on the cell behavior during a neoplastic transformation or during other pathological processes.

2.6. *'Hic sunt leones!'*

On old adventurers maps sometimes the inscription was found 'Hic sunt leones!', meaning that lions occur in that territory, or the territory is undisclosed and dangers are hiding in it. With respect to the pathobiology of cancer, many questions remain to be solved and await their inclusion into standard therapy. For instance cancer cells escape immune surveillance but many aspects of cancer biology also escape from being incorporated into disease modeling.

Recent and future progress in disease modeling has focused on the following aspects:

1) Factors that cause a (group of) cells to escape from proliferation control;

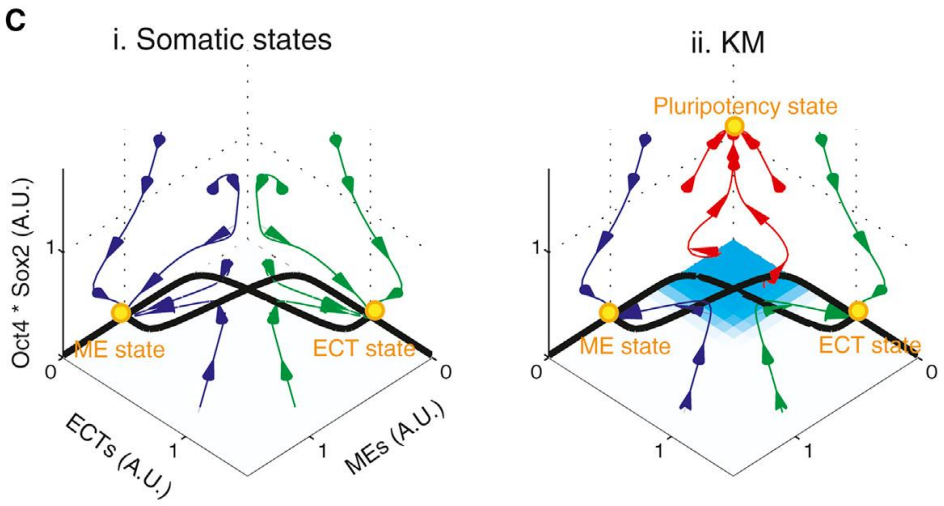

Figure 3. Cell-state landscapes, presented by Shu et al. (2013). Cell-state landscapes are obtained from the density of trajectories, (i) for the 'somatic state' and (ii) for the 'KM state': according to Polo et al. (2012), overexpression of KLF4 and c-Myc (KM) may act as a driving force that reduces 'reprogramming barriers'. The blue color in the KM cell-state landscape of Shu et al. (2013) indicates deep 'attractors' near which the cell states tend to stay. The x-axis represents the difference between the two groups of lineage inducers, and the y-axis indicates the 'pluripotency' of the cell. (© 2013, Elsevier Inc., Cell, Vol. 153, p. 972)

2) Factors that cause a (group of) cells to escape from apoptosis;

3) Factors that cause a (group of) cells to escape from immune surveillance;

4) Factors that cause a (group of) cells to escape from uptake of (toxic) chemicals.

Our challenge is to incorporate these metaphorical 'lions' into a comprehensive, multiscale model, or to define the landscape architecture in which the latter events take place.

3. Defining the Architecture for Multiscale Modeling

3.1. Symmetry and symmetry breaking

In his thought-provoking monograph Life's other secret Ian Stewart (1998) proposes the following rhetorical question: "*Are there mathematical laws for the behavior of cells? How did cells get together to form what we call 'higher' organisms? Are mathematical laws behind those, too?*" (Stewart, 1998, p. 75). And the answer to that rhetorical question soon follows (referring to the function of tubulin in the cytoskeleton, the centrosome and its role in cell division): "*It's ironic: The very kinds of behavior that make cells seem so organic and nonmathematical actually have a mathematical origin. Clearly, we have the wrong idea about the capabilities of mathematics.*" (Stewart, ibidem).

Theoretically, like in the cell-free model (see SECTION **2.2**), if the state of a system of N molecular species can be described by a number (M) differential equations (ODEs), the 'symmetry' of this system can be specified in terms of a group of transformations of the variables that "in some sense preserves the structure of the equation – in particular,

its solutions."(Golubitsky and Stewart, 2002, p.6). Mathematically, this symmetry group refers to the notion of a Lie group, namely the 'symmetric group' (S_M) consisting of all permutations on M symbols. This Lie group can also be represented as a differentiable manifold (Hawkins, 2000). The symmetries of solutions of the system, however, do not equal the symmetries of the system in general (Golubitsky and Stewart, p. 8 a.f.). This is especially important for the notion of bifurcation, for bifurcation branches will correspond to symmetry-breaking solutions. Moreover, steady-state bifurcations will depend on the (zero) eigenvalue of the linearized equation, or, the generalized eigenspace corresponding to zero is the kernel of the linearized equation (Golubitsky and Stewart, 2002).

In experimental gene expression studies however, it is not very easy to obtain differentiable equations of time-dependent expression data. For most sampling techniques (like PCR) allow for estimating the amount of expression at a given time-point but not the dynamic expression patterns. Nevertheless, we may retain the theoretical notion of symmetry-breaking bifurcation branches (referring to a group of spatio-temporal symmetries) (Golubitsky and Stewart, 2002) in cellular development and reprogramming.

3.2. Landscape Architecture for Cell Development and Reprogramming

Recently, Shu *et al.* (2013) developed a so-called coupled pluripotency module for the self-activation of the Oct4 and Sox2 genes, and for the mutual antagonism between the mesendodermal (ME) and ectodermal (ECT) lineage specifications (**FIG. 2**). Moreover, they designed a cell-state landscape obtained from the trajectories of cell differentiation states (**FIG. 3**). Their so-called 'see-saw' model incorporates a combination of activating and inhibitory patterns and elaborates on the notion of 'trajectory density', also known from signaling network studies (Ma'ayan *et al.*, 2005; Bromberg *et al.*, 2008).

Based on a number of existing data bases of mammalian protein-protein interactions, Avi Ma'ayan *et al.* (2005) developed a (neuronal) signaling network, in which the density of information processing (DIP) is defined as follows:

$$DIP_i = \left(\frac{M_i - M_{i-1}}{L_i - L_{i-1}} \right) GC_i \qquad \textbf{(11)}$$

with M_i the number of feedback loops at step i, L_i the number of links at step i and GC_i the grid coefficient, representing the interconnectedness for a given subnetwork at step i (Ma'ayan *et al.*, 2005).

Similar to this DIP, the trajectory density is used by Shu *et al.* (2013) to define 'attractors' near which the cells tend to stay (**FIG. 3**). On the other hand, in the work of Polo *et al.* (2012) cellular reprogramming intermediates are defined on the basis of gene clustering, studied with PCA (see SECTIONS **2.4-2.5**). Susceptibility to reprogramming, or to change in general, should be inversely proportional to this interconnectedness: it may depend on the sensitivity of the network node to extracellular and/or environmental cues, and also to their proximity to inherent high risk genes for cancer development (defined as a distance function on the interaction grid). An example of the identification of genomic regions with frequent alterations in cancer, based on high-throughput anal-

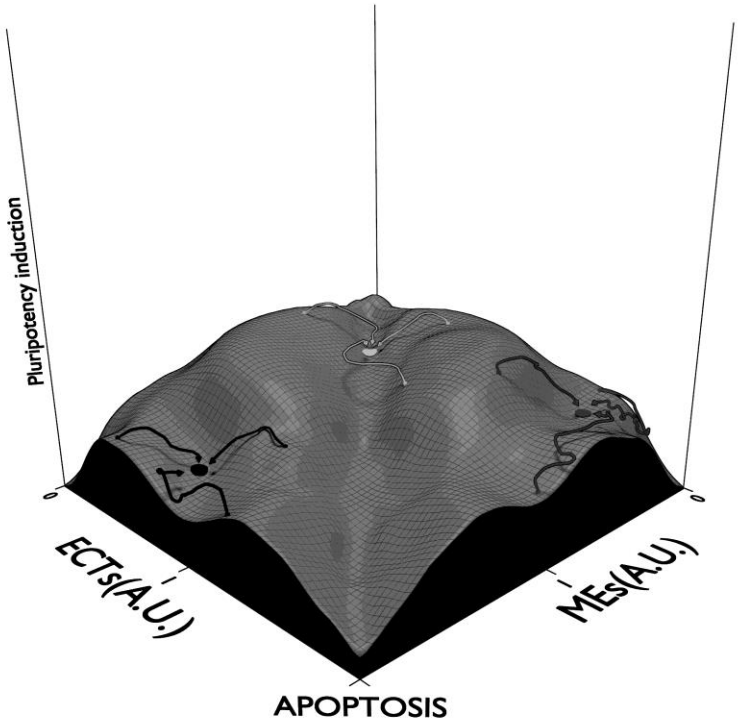

Figure 4. Artist impression of cell-state landscape, in analogy with the work of Shu *et al.* (2013). A two-dimensional cell-state potential function is drawn, with reference to the ECT and ME lineage inducing factors as studied by Polo *et al.* (2012) and Shu *et al.* (2013). Also we added 'apoptosis' as the final developmental state. Trajectories and presumed 'attractors' may reflect the local behaviour of cells within a chosen experimental set-up and cellular environment.

ysis of somatic copy-number alterations (SCNAs), was published by a consortium of researchers (Beroukhim *et al.*, 2010). It was found that cancer cells contained amplifications surrounding specific gene families, especially the *MCL1* and *BCL2L1* anti-apoptotic genes (family of apoptosis regulators). Sensitivity of epithelial cells to become migratory cells, are for instance related to transformation of proteins of the E-cadherin complex, that constitutes adherens junctions between epithelial cells. Further modeling and experimental studies will be necessary, however, to incorporate other decisive steps in cell development, similar to the well-known induction of apoptosis in relation to the inhibition of tumor growth (Shi *et al.*, 2005).

These challenges are enticing to draw a comprehensive cell-state landscape reflecting the multiscale propensities of cells during normal development or neoplasm formation. The cell-state landscape has the characteristics of a potential function $\Phi(\eta, \zeta)$, e.g. using the arbitrary units for MEs and ECTs used by Shu *et al.* (2013) (**FIG. 4**). However, this is only one of a set of possible representations, namely a 2D-projection of a hyperbody representing all possible cell states on the plane defined by two arbitrary axes and according to a third axis (pluripotency potential). Also other projections of this multi-dimensional cell state manifold are conceivable, following the use of specific

cell lineage inducers (**FIG. 3, 4**).

The trajectories on these curved surfaces are related to the probability of changing from one state $M\ (\eta_i\ ,\ \zeta_j\ )$ to another state $N\ (\eta_{i+\Delta},\ \zeta_{j+\Delta}\ )$, and have the dimension of a probability squared ρ^2 (Allaerts, 2008). On the other hand, the surface also represents a potential function, for which the integration from M to N corresponds to the curvilinear integral of the trajectory

$$\int_M^N \eta d\eta + \zeta d\zeta = \int_M^N d\Phi(\eta,\zeta) \qquad \textit{(12)}$$

According to integration rules for curves on a differentiable manifold, we obtain:

$$\int_M^N d\Phi(\eta,\zeta) = \Phi(N) - \Phi(M) \approx \rho_\Delta^2(\eta,\zeta) \qquad (13)$$

Following the stochastic rules (see SECTION **2.2**) for modeling an aleatory 2D-variable with reference to some local point maximum, the probability density in a local environment of the states M or N corresponds to a 2D-normal distribution function:

$$f(\eta,\zeta) = \frac{1}{2\pi\sigma_\eta\sigma_\zeta} e^{-\frac{\eta^2}{2\sigma_\eta^2}-\frac{\zeta^2}{2\sigma_\zeta^2}} \qquad \textit{(14)}$$

The cell-state potential function depends on the balance between $\Sigma\ DIPi$ [eq. ***(11)***] for a cell at a particular cell state and the susceptibility to change for a particular cellular environment (*e.g.*, expression of apoptotic genes or genes for cell adhesion molecules). The trade-off between density of information and susceptibility to change, may correspond to bifurcation patterns (see SECTION **3.1**) in specific projections of the multidimensional cell-state manifold.

As shown in **FIG. 4**, a typical cell-state landscape encompasses both global patterns for cell development as well as local irregularities that reinforce the local behavior of cells in a perceived (or semi-) deterministic pattern. In the future, refinements of this type of cell-state modeling may very well consist of defining the local ridges and grooves for a particular cellular transformation landscape, based on experimental (high-throughput) transcriptome and molecular interaction data.

References

Allaerts W [2008] Towards the development of annotation grids for cell lineage studies: I. Study of paradigm shifts. Med Hypotheses Res 5: 37-51.

Allaerts W [2011] New cellular concepts in Neuroendocrinology: The anterior pituitary case. Curr Trends Endocrinol 5: 19-28.

Beroukhim R, Mermel CH, Porter D, *et al.* [2010] The landscape of somatic copy-number alteration across human cancers. Nature 463: 899-905.

Brinkmann HW and Klotz EA [1971] Linear algebra and analytic geometry. Reading, Menlo Park, London: Addison-Wesley Publishing Company, 535 pp.

Brognard J and Dennis PA [2002] Variable apoptotic response of NSCLC cells to inhibition of the MEK/ERK pathway by small molecules or dominant negative mutants. Cell Death Differ 9: 893-904.

Bromberg KD, Ma'ayan A, Neves SR and Iyengar R [2008] Design logic of a cannabinoid receptor signaling network that triggers neurite outgrowth. Science 320: 903-909.

Dasen JS, O'Connell SM, Flynn SE, Treier M, Gleiberman AS, Szeto DP, Hooshmand F, Aggarwal AK and Rosenfeld MG [1999] Reciprocal interactions of Pit1 and GATA2 mediate signaling gradient-induced determination of pituitary cell types. Cell 97: 587-598.

Duncan OD [1975] An Introduction to Structural Equation Models. New York: Academic Press.

El-Samad H and Khammash M [2006] Regulated Degradation is a Mechanism for Suppressing Stochastic Fluctuations in Gene Regulatory Networks. Biophysical J 90: 3749-3761.

El-Samad H, Kurata H, Doyle JC, Gross CA and Khammash M [2005] Surviving heat shock: Control strategies for robustness and performance. Proc Natl Acad Sci USA 102: 2736-2741.

El-Samad H, Prajna S, Papachristodoulou A, Doyle J and Khammash M [2006] Advanced Methods and Algorithms for Biological Networks Analysis. Proceedings of the IEEE 94: 832-853.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES [1999] Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286: 531-537.

Golubitsky M and Stewart I [2002] The symmetry perspective. From equilibrium to chaos in phase space and physical space. Basel, Boston, Berlin: Birkhäuser Verlag.

Gould PR [1967] On the geographical interpretation of eigenvalues. Trans Inst Brit Geogr 42: 53-86.

Hawkins T [2000] Emergence of the theory of Lie groups. Sources and studies in the history of mathematics and physical sciences. Berlin, New York: Springer Verlag.

Hess KR, Varadhachary GR, Taylor SH, Wei W, Raber MN, Lenzi R and Abbruzzese JL [2006] Metastatic patterns in adenocarcinoma. Cancer 106: 1624-1633.

Johnston RJ [1980] Multivariate Statistical Analysis in Geography. A Primer on the General Linear Model. London, New York: Longman, 280 pp.

Kaiser HF [1958] The varimax criterion for analytic rotation in factor analysis. Psychometrika 23: 187-200.

Koopman G, Reutelingsperger CP, Kuijten GAM, Keehnen RM, Pals ST and van Oers MH [1994] Annexin V for flow cytometric detection of phosphatidylserine expression on B cells undergoing apoptosis. Blood 84: 1415-1420.

Ma'ayan A, Jenkins SL, Neves S, Hasseldine A, Grace E, Dubin-Thaler B, Eungdamrong NJ, Weng G, Ram PT, Rice JJ, Kershenbaum A, Stolovitzky GA, Blitzer RD and Iyengar R [2005] Formation of regulatory patterns during signal propagation in a mammalian cellular network. Science 309, 1078-1083.

Mansury Y and Deisboeck TS [2003] The impact of "search precision" in an agent-based tumor model. J Theor Biol 224: 325-337.

Mercier F and Hatton GI [2001] Connexin 26 and basic fibroblast growth factor are expressed primarily in the subpial and subependymal layers in adult brain parenchyma: roles in stem cell proliferation and morphological plasticity? J Comp Neurol 431: 88-104.

Mouneimne G, Soon L, DesMarais V, Sidani M, Song X, Yip SC, Ghosh M, Eddy R, Backer JM and Condeelis J [2004] Phospholipase C and cofilin are required for carcinoma cell directionality in response to EGF stimulation. J Cell Biol 166: 697-708.

Munsky B and Khammash M [2006] The finite state projection algorithm for the solution of the chemical master equation. J Chem Phys 124: 044104.

Nguyen DX [2011] Tracing the origins of metastasis. Journal of Pathology 223: 195-204.

Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alaei S, Cloutier J, Bar-Nur O, Cheloufi S, Stadtfeld M, Figueroa ME, Robinton D, Natesan S, Melnick A, Zhu J, Ramaswamy S and Hochedlinger K [2012] A molecular roadmap of reprogramming somatic cells into iPS cells. Cell 151, 1617-1632.

Shi J, Zheng D, Liu Y, Sham MH, Tam P, Farzaneh F and Xu R [2005] Overexpression of soluble TRAIL induces apoptosis in human lung adenocarcinoma and inhibits growth of tumor xenografts in nude mice. Cancer Res 65: 1687-1692.

Shu J, Wu C, Wu Y, Li Z, Shao S, Zhao W, Tang X, Yang H, Shen L, Zuo X, Yang W, Shi Y, Chi X, Zhang H, Gao G, Shu Y, Yuan K, He W, Tang C, Zhao Y and Deng H [2013] Induction of pluripotency in mouse somatic cells with lineage specifiers. Cell 153, 963-975.

Stewart I [1998] Life's other secret. The new mathematics of the living world. New York, Brisbane, Singapore, Toronto: John Wiley & Sons, Inc., 285 pp.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K and Yamanaka S [2007] Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell 131: 861-872.

The Cancer Genome Atlas Research Network [2008] Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455: 1061-1068.

Tindemans S [2009] Biomolecular design elements. Cortical microtubules and DNA- coated colloids (p. 78). PhD thesis, Wageningen University.

Vicent S, Garayoa M, Lopez-Picazo JM, Lozano MD, Toledo G, Thunnissen FB, Manzano RG and Montuenga LM [2004] Mitogen-activated protein kinase phosphatase-1 is overexpressed in non-small cell lung cancer and is an independent predictor of outcome in patients. Clin Cancer Res 10: 3639-3649.

Wang Z, Bordas V and Deisbroeck TS [2011] Identification of critical molecular components in a multiscale cancer model based on the integration of Monte Carlo, resampling, and ANOVA. Frontiers in Physiology 2: 35.

Wang Z, Zhang L, Sagotsky J and Deisboeck TS [2007] Simulating non-small cell lung cancer with a multiscale agent-based model. Theoretical Biology and Medical Modelling (BioMed Central) 4: 50.

Yamasaki H and Nauss CC [1996] Role of connexin genes in growth control. Carcinogenesis 17: 1199-1213.

Zhang HX, Dempsey WP Jr and Goutsias J [2009] Probabilistic sensitivity analysis of biochemical reaction systems. J Chem Phys 131: 094101.