Abstract

Epigenetic Tracking is a biological model, originally conceived to study embryonic development. Computer simulations proved the capacity of the model to generate complex 3-dimensional cellular structures, and the potential to reproduce the complexity typical of living beings. The most distinctive feature of this model is the presence in the body of a homogeneous distribution of stem cells, which are dynamically and continuously created during development from non-stem cells and reside in niches. Embryonic stem cells orchestrate early developmental events, adult stem cells direct late developmental and regeneration events, ageing stem cells cause ageing and cancer stem cells are responsible for cancer growth. The conceptual backbone provided by Epigenetic Tracking brings together a wide range of biological phenomena: for this reason, we think it can be proposed as a general model for multicellular biology. Despite, or perhaps due to its theoretical origin, the model allowed us to make predictions relevant to very diverse fields of biology, such as transposable elements, and cancer-related patterns of gene mutations. This paper contains a summary of the model and its implications.

1 Introduction

This work is concerned with a biological model called Epigenetic Tracking (ET), described in [6, 7, 8, 9, 10, 11, 12, 13, 14]. While the study of life sciences often relies on a bottom-up method, that tries to infer general rules starting from genes and proteins, our work is informed by a top-down approach, that proceeds from high-level abstractions towards a proposal for biological, low-level molecular processes. As a consequence our model may contain ingredients not necessarily adherent to current knowledge, but which can become a suggestion for biologists to look into new, previously unexplored directions.

The capacity of the model to generate complex cellular structures was proved through computer simulations. If we interpret such structures as a metaphor for biological organisms, we can see in this endeavour the potential to reproduce the complexity typical of living beings. Furthermore, ET is able to interpret with a unified framework a wide array of biological
phenomena, such as development, the presence of junk DNA, ageing and cancer: for this reason, we think it can be proposed as a general model for multicellular biology. This paper is divided into six sections: this (first) section is the introduction; section 2 describes ET as a model of development; section 3 is dedicated to the evo-devo process; section 4 deals with ageing; section 5 explores the topic of cancer; section 6 draws the conclusions and outlines future directions.

2 Development

HOX genes are a family of regulatory genes, characterised by the presence of a DNA sequence called homeobox. Their gene products are transcription factors that control the activity of other, functionally related genes in the course of development. In early embryogenesis, HOX gene regulation is achieved through the maternal factors present in the fertilised egg, gap and pair-rule genes [26]. In subsequent phases of development this role is taken over by morphogens, a class of compounds (e.g., BMP, Hedgehog, Nodal, Notch, Wnt) known to play a role in tissue patterning, and involved in cancer formation. During development, they spread from localised sources and form concentration gradients, which can be used by cells to induce the expression of specific genes, through a characteristic “morphogenetic code” [16]. Each morphogen is processed by a dedicated pathway, that starts when a ligand binds a receptor on the cell surface, and ends in the nucleus, where the induced signal participates in the regulation of target genes.

In biological development a pivotal role is known to be played by stem cells, a class of cells found in most multicellular organisms. Embryonic stem cells (found in the inner cell mass of the blastocyst) are totipotent cells, able to differentiate into all cell types of the body. Adult stem cells are pluripotent undifferentiated cells found throughout the body after embryonic development. Unlike embryonic stem cells, adult stem cells can only form a limited set of cell types and function to replenish dying cells and regenerate damaged tissues. The question of when and how adult stem cells are formed remains an open issue in biology.

In ET development starts from a single cell and unfolds in a predefined number of developmental stages, counted by a global clock shared by all cells. Artificial bodies are composed of two categories of cells: normal cells and stem cells. Cells (Fig. 1) have a variable called mobile code, which can take different values in different cells and represents the main source of differentiation during development. A given mobile code can be shared by many normal cells, but each stem cell has a distinct mobile code.

The mobile code can be interpreted in biology as the set of master transcription factors present in a cell, which are responsible for the activation of batteries of other genes, and ultimately determine the establishment of cell type. In nature the mobile code is hypothesised to be composed of HOX proteins. The clock represents a further transcription factor, carrying a temporal condition for gene activation.

In ET cells have an associated genome, composed of different elements: morphogen processing modules, the HOX module and effector genes (Fig. 1). For simplicity we represented only four morphogen processing modules, named as the corresponding biological pathways:
**Figure 1:** Structure of a cell in ET. The mobile code can be interpreted in biology as a set of master transcription factors. It is structured as a sequence of numbers and indicated with a letter. The genome is composed of different elements: the morphogen processing modules, the HOX module and the effector genes. For simplicity only four morphogen processing modules are represented: the BMP module, the Nodal module, the Notch module and the Wnt module. Effector genes are divided in groups of elements which are co-regulated: for this reason each group is associated to a variable called regulatory set and a variable called timer.

**Figure 2:** Change events in ET. The figure shows an example of a proliferation event triggered by the stem cell with mobile code K at developmental stage 6 (when the clock value is 6). As a result one gene group is activated and the outcome is a proliferation event. Most new cells become terminally differentiated, while other cells revert to the stem state.
**Figure 3:** Biological mobile code, clock, regulatory set and timer. The figure shows the components of a biological mobile code binding to the components of a biological regulatory set and a biological clock binding to biological timers. As a result, two biological genes are activated (only two genes are shown for simplicity).

**Figure 4:** Set of stem cells created during development. Stem cells marked in green become active during development. As a result, they create normal cells and induce their differentiation. In parallel, other stem cells are created.
Stem cells direct the developmental process. When a stem cell (that we call *mother cell*) reaches a particular state at a particular time, a specific *change event* is orchestrated. As an abstraction for this process, we use the match between the mobile code and the regulatory set of a gene group, and between the clock and the timer (Fig. 2).

Morphogen processing modules and the HOX module correspond to biological morphogen processing pathways and to HOX genes, respectively. Each component of the mobile code / regulatory set can be interpreted as a transcription factor / regulatory locus, while the switch may correspond in biology to a methylation mark. The match between the mobile code and the regulatory set corresponds in biological terms to the binding of master transcription factors to gene regulatory regions (Fig. 3). Analogous considerations apply to the clock and the timer.

The change event that takes place once the match occurs is encoded in the gene group, and can take the form of either a large-scale apoptosis (death of a large number of cells in the volume around a stem cell), or a proliferation event, which corresponds to several rounds of asymmetric cell divisions (i.e. the stem cell produces another identical stem cell and a normal cell). A proliferation event initially produces a large number of normal cells in the volume (Fig. 2, panel A). While most normal cells become terminally differentiated cells, some other cells, instead of proceeding on their differentiation path, revert to stem cells (Fig. 2, panel B). Hence, in ET the normal-stem transition is a two-way street.

The normal-stem biunivocal transition may entail a fundamental principle present also in nature, that induces cellular systems to self-organise in a hierarchical fashion. We think of this *biological hierarchical principle* as a pervasive mechanism underlying all manifestations of cellular life, in both physiological and pathological conditions.

The new stem cells generated from normal cells receive a new and unique mobile code. If in the genome a gene group exists whose regulatory set matches the mobile code of a new stem cell, this cell can become the centre of another event in a subsequent stage, and development can move ahead. This cycle, repeated over and over again, represents the core of the morphogenetic process.

The set of stem cells and mobile codes generated during development has a hierarchical, or tree, structure (Fig. 4). This descends from the fact that each stem cell originates from the conversion of a normal cell, which has a single mother. This property is true for single body parts and for the whole body.

This feature appears consistent with the information we have on the set of biological stem cells involved in the generation of particular organs or systems, such as the hematopoietic...
Figure 5: Morphogens in ET. The panel on the left shows a hypothetical distribution of stem cells. Some of these stem cells (dashed circles) emit morphogens, of a finite number of types, which diffuse across the space. On the right, the different combinations of concentration values of the morphogens partition the space into regions (marked each with a different colour). In each such region a new stem cell is created from a normal cell (red circles).

system [27]. Our model provides also a means to bridge the conceptual gap between embryonic and adult stem cells: embryonic stem cells correspond to elements in the tree near the root, while adult stem cells correspond to the leaves of the tree. It also explains how adult stem cells are generated, with the mechanism of conversion from normal cells.

In ET the process of normal-stem conversion is achieved through the presence of simulated morphogens, with the following mechanism. Some stem cells (Fig. 5) become sources of specific morphogens (belonging to a number of distinct types). These signals partition the space into regions, each characterised by a distinct combination of concentration values. For example, all cells in a given region “sense” one morphogen with concentration 7, another morphogen with concentration 4, etc. (concentration values are integer numbers which represent categories or ranges of actual concentration values). In each region a normal cell is selected and turned into a stem cell.

To each new stem cell a new and unique mobile code is assigned. When the stem cell is created, it has the same mobile code as the normal cell from which it was obtained. The assignment of a new mobile code is achieved through the processing of the signals received by morphogen processing modules. The outcome of this network is fed to the Hox module, whose product is a number which is added to the existing numbers already present, to produce the new mobile code for the stem cell (Fig. 6). Figure 7 shows an example of development for a hypothetical “organ”, which alternates the occurrence of change events and the creation of new stem cells.

According to our hypothesis, in biology each morphogen processing pathway senses the concentration of a specific morphogen, and sends a signal into the nucleus. Here the interplay of signals conveyed by all pathways result in the transcription of a different HOX gene. For example, if the combination of concentrations is [2130] (e.g. one morphogen is received with concentration 2, another morphogen with concentration 1, etc., HOX gene X is transcribed. If the combination of concentrations is [3110], HOX gene Y is transcribed. The particular com-
**Figure 6:** Normal-stem transition and generation of a new mobile code in ET. Morphogens, released by some stem cells (indicated with dashed circles in the corners), are processed by morphogen processing modules and by the HOX module. The output is a new number (corresponding in nature to a new HOX protein), which is added to the repertoire of numbers already present in the cell.

**Figure 7:** Stem cells drive organogenesis in ET. The figure shows the development of a hypothetical organ in two stages, carried out by five change events (stem cells are marked in yellow, normal cells just created -plastic- in orange, normal differentiated cells are shown with different colours, each corresponding to a particular differentiation state). The right part of the figure reports the mobile codes generated during development, which are organised in a tree-like structure. Each stage is shown in two panels, A and B. Panel A shows the occurrence of one or more change events, panel B shows the creation of new stem cells. This process is based on morphogens emitted by a subset of stem cells, which partition the space in regions. One normal cell in each region is turned into a stem cell, and obtains a distinct mobile code.
Figure 8: Self-generation and self-regeneration of a lizard-shaped structure consisting of 500000 cells in ET. Development unfolds from a single cell (circle) in 12 developmental stages (frames 1-12). After stage 12, the head and the tail are cut off (arrows). After the debris is removed, the structure is completely regenerated.

Combinations of morphogen concentrations necessary for mobile code generation may correspond to the set of signals required to maintain stem cells in niches [28].

The proposed framework allows also to model cellular regeneration. Our concept of regeneration is based on the idea to recreate exactly the same conditions as during development. Stem cells, after their activation during development, are kept in a deactivated state (so that the associated event is not triggered again). Once a damage is detected in a given body part, the stem cells activated to produce that body part during development are reactivated during regeneration, leading to the same sequence of events. The regeneration is perfect, as our simulations confirmed (Fig. 8).

3 Evo-devo and transposable elements

The model of development described has been coupled with a genetic algorithm, able to simulate Darwinian evolution. The genetic algorithm evolves a population of individuals (each encoded in an artificial genome) for a number of generations. At each generation, all individuals develop independently from the zygote stage to the final phenotype, whose proximity to a predefined target is employed as a fitness measure. This operation is repeated for all individuals, so that eventually each individual is assigned a fitness value. Based on this value the genomes of the individuals are selected and randomly mutated, to produce a new population. This cycle is repeated until a satisfactory level of fitness is reached. The coupling of the
model of development and the genetic algorithm gives origin to an evo-devo process, which was proved able through computer simulations to “devo-evolve” 3-dimensional structures of unprecedented complexity (see examples in Fig. 9).

In ET most stem cells produced during development do not orchestrate any events (are inactive), because in the genome there is no gene group whose regulatory set can match their mobile code (upper left part of Fig. 10). The probability that a suitable regulatory set (which is encoded as a long sequence of numbers) emerges in the genome simply through mutations and recombinations is very low. A countermeasure consists in “suggesting” to the genetic algorithm to put in the genome some regulatory sets which are guaranteed to match. This idea is implemented in a procedure called germline penetration, which builds regulatory sets able to match mobile codes generated during the development of the structure. These regulatory sets are then associated to gene groups contained in a special copy of the genome called “germline” genome (upper right part of Fig. 10) which, after reproduction, is destined to become the (“somatic”) genome of the individuals of the next generation.

Once evolution is provided with a “good” regulatory set, i.e. one guaranteed to match an existing mobile code, it has to optimise the coding parts of the associated genes, a process that can take several generations. When the optimisation of the coding parts is completed, the new gene group can be activated and carry out another change event (lower part of Fig. 10). New mobile codes are generated as a result of the new change events: regulatory sets able to match such new mobile codes are again transferred to the germline genome, and the whole cycle repeats itself. The gene groups associated to the penetrated regulatory sets are initially set as inactive: otherwise, the genes would become active with a non-optimised coding part, causing a disruption in development (and an abrupt decrease in the individual’s fitness). Their activation is obtained through a subsequent genomic mutation of the switch. Therefore,
Figure 10: Germline penetration of new regulatory sets in ET. In the upper part, in a stem cell the genome is not able to “respond” to the mobile code M with a corresponding event. A new regulatory set able to respond is created in the stem cell. The new regulatory set leaves the stem cell and reaches the germline, where it is incorporated in the genome. In a subsequent generation, when a cell has mobile code M, its genome (inherited from the penetrated germline) can respond with a specific gene group. The result is a proliferation event.

According to ET, evolution is essentially Lamarckian for regulatory sequences and Darwinian for coding sequences. One point is worth stressing.

Germline penetration is essential in ET for the evolution of complex shapes: in our in silico experiments, if germline penetration is disabled, the evolutionary process practically grinds to a halt. The central role played by germline penetration in our model lead us to hypothesise the existence of a similar process also in biological organisms. For the implementation of a biological germline penetration, we need genetic elements able to build new regulatory sequences in somatic cells, and susceptible of being transferred to germline cells.

Transposable elements, or transposons [24], are DNA sequences that can move around to different positions in the genome. During this process, they can cause mutations, chromosomal rearrangements and lead to an increase in genome size. Despite representing a large genomic fraction (30-40% in mammals), transposons have no clear biological function, and have therefore been labelled as “junk DNA”. Transposable elements are active during devel-
velopment, and may bring diversity in somatic cells having the same genome [2]. A transposomal activity was also observed in germline cells, with possible evolutionary implications. “Waves" of diffusion of transposable elements in a genome appear temporally associated with major evolutionary changes to the species [25], suggesting a causal link between the two events.

Based on these considerations, we can imagine a mechanism for the implementation of a biological germline penetration. In a stem cell unable to produce a change event (because the set of master transcription factors present in the cell cannot bind to corresponding regulatory sequences), transposable elements would become mobilised, and start “jumping" around the genome. This process could lead to the creation of many new regulatory sequences, hugely increasing the chances that a lucky combination of such sequences be able to match the existing transcription factors. We hypothesise that, when a match occurs, the transposable elements which helped to build the new sequences leave the cell and, through the bloodstream, make their way to germline cells. Here they would insert themselves into the genome, thus allowing the delivery of a successful innovation to the next generation.

The mechanism described has interesting evolutionary implications. In ET, whenever a stem cell proliferates, a wave of new stem cells, each with its own mobile code, is created in the body of the (new) species. The action of germline penetration translates this wave of new mobile codes in somatic cells into a corresponding wave of new regulatory sets spreading in the germline genome and subsequently in the somatic genome of future generations. Such events during artificial evolution coincide with moments in which major changes occur to the evolving species, causing new body parts or features to appear.

In biological terms, this means that the spreading in the genome of waves of new sets of transposable elements in the course of evolution corresponds to moments in which new branches (new species) are generated in the “tree of life”. Such predictions, made with our model on purely theoretical grounds, appear to be confirmed by experimental evidence [25]. These observations suggest that the spreading of new transposon families in the genome and the occurrence of major changes in the relevant lineage are simultaneous events. This seems to hint that the colonisation of the genome by the transposons is the driving force behind the change in the lineage. Our interpretation of this phenomenon is different. In fact, according to ET, the spread of new transposon families in the genome is an event which comes immediately (in evolutionary terms) after the change, not before. The confirmation of this prediction, made possible by techniques to estimate the age of DNA sequences more sensitive than those currently employed, would represent a clear indication in favour of this model.

In conclusion, our model suggests a biological role for transposable elements which, as we have seen, account for up to 40% of the genomic content in mammals. We argue that the presence of trasposable elements is not an artifact, but an essential biological feature. In the ET framework, for any individual, the set of mobile codes generated during development can be divided into i) mobile codes that activate a gene group during development and ii) mobile codes that do not activate any gene group during development. Indeed, the ET machine cannot avoid generating inactive mobile codes. To prevent this effect we should reduce the density of stem cells, but this would also reduce the effectiveness of the morphogenetic pro-
cess. Therefore, the presence of a certain amount of inactive mobile codes is unavoidable. On this material intervenes germline penetration, converting inactive mobile codes into a corresponding number of inactive regulatory sets in the genome (the penetrated regulatory sets are inactive because the associated switches are initially set to OFF). Therefore, the presence of inactive information in both the set of mobile codes and the genome is a fact which is inescapably connected to the core of the ET machine, a requirement essential to its evolvability.

4 Ageing

Ageing is a universal phenomenon, for which many theories exist. Stochastic theories (e.g. the “free-radical theory” [15]) blame damage induced by environmental factors, and accumulating over time, as the cause of ageing. According to programmed theories (e.g. the “ageing-clock theory” [5]), ageing is driven by genetic instructions, which can to a certain extent be modulated by environmental conditions. Evolutionary theories (e.g. the “disposable soma theory” [19]) see evolution as the main force shaping the ageing profile of different species.

A measure that was proved effective in reducing the rate of ageing in many species is dietary restriction [23]. Many genes have been shown to influence the life span in some species; most of these genes belong to the insulin / IGF1-like signalling pathway, to the class of sirtuins, and to the target of rapamycin (TOR) pathway [18]. This last pathway is thought to mediate the effects of dietary restriction at the cellular level. Overall, it is safe to say that both genetic and environmental factors concur to cause ageing, but the balance between these factors appears to change in the course of life: the first seem to prevail in the young, the second become dominant in the old. We will now propose a possible explanation for this phenomenon.

In ET, for a given individual development unfolds in N developmental stages; at the end of it the individual’s fitness is evaluated, and used for reproduction in the genetic algorithm. The moment of fitness evaluation, that in nature may correspond to the moment of reproduction, in most of our experiments has coincided with the end of the simulation. On the other hand, we can imagine to let the global clock tick on and see what happens in the period after fitness evaluation. The distinction between the periods before and after fitness evaluation can be thought to correspond to the biological periods of development (say, until 25 years of age in humans) and ageing (from 25 years of age onwards).

At the end of an individual’s development, many stem cells are present in the body of the individual. Some of these stem cells have been activated during development (and shaped the individual’s body), some (the vast majority in our simulations) have not (Fig. 11). These stem cells may contain gene groups bound to be activated in the ageing period, after the moment of fitness evaluation (the moment of activation is determined by their timer value, \( \geq 25 \)) when, by definition, they are not affecting the fitness value. For this reason, their coding parts have not been optimised by evolution and the associated events will tend to have a “pseudo-random” character, i.e. they will look random, even though they are not (they are encoded in the genome and, as such, entirely under genetic control and deterministic).

The consequence of the “pseudo-randomness” of these events is that their effects on the
Figure 11: Stem cells present in ET artificial bodies at the end of development. Some of these stem cells have been activated during development, some (the vast majority) have not. The timer value of the gene group which is going to be activated in each stem cell is indicated in the circles (numbers refer to age in humans expressed in years).

Figure 12: In the upper panel: example of ageing in ET for a “face”. On the left the period of development: the structure grows from a single cell to the mature phenotype at age 25 (red frame), when fitness is evaluated. On the right the period of ageing: the quality of the structure deteriorates steadily under the action of non-optimised gene groups. In the mid panel: the evolutionary pressure acting upon a gene group (represented by the darkness level) is an inversely proportional function of the group timer value. Gene groups with timer values lower than 25 years are subject to a high evolutionary pressure, gene groups with timer values greater than 25 years are subject to a steadily decreasing pressure, until their effects become pseudo-random. In the lower panel: the level of pseudo-randomness (represented by the darkness level) of a gene group’s coding part increases with the gene timer value.
overall individual's health are more likely to be detrimental than beneficial. Here the term “health” is used to indicate the individual “physical” condition, while the (reproductive) “fitness” corresponds to the chance that the individual’s genes will survive in the future genetic pool of the population. As a result, in the ageing period, the individual's health will tend to progressively decrease over time, under the action of such pseudo-random events. Therefore, ET sees ageing as a continuation of development, driven by non-optimised genes activated in specific stem cells after the age of reproduction. Our postulate is that these considerations apply also to biological ageing.

The hypothesis, according to which the evolutionary pressure acting on a gene group is null if it is activated after reproduction, is rather simplistic. In nature, an individual's reproductive fitness depends also on events manifesting themselves after reproduction, as also these affect the survival chances of its progeny. More realistically, the effect of an event on the fitness (and hence the evolutionary pressure acting upon the corresponding gene group) will tend to decrease as the age of its manifestation increases, rather than vanishing immediately after reproduction. Figure 12 shows a simulation of ageing for a “face”, which incorporates this last observation.

The theory proposed provides a synthesis of all three categories of ageing theories. It has a solid evolutionary background, and explains why the experimental evidence has a deterministic connotation at young age and turns stochastic with age progression. It has also the advantage to interpret development and ageing with a common theoretical framework. In biology, besides proliferation and apoptosis, we can envision the existence of other types of events. Another “biological change event” could induce epigenetic changes in the gene regulatory networks of affected cells. This, in turn, would change the biochemistry of the relevant tissue and the function of the organ.

Another merit of our proposal is the involvement of stem cells in the picture. By analogy with “cancer stem cells”, we may call the stem cells involved in the ageing process “ageing stem cells”. There is empirical evidence of an age-related decline in the functionality of adult stem cells [22]. Our hypothesis is that this functional decline is determined by epigenetic changes induced in stem cells by biological change events, bound to occur at precise moments. Such biological change events may occur both in stem cells activated during embryonic development, and reactivated in the post-developmental period for tissue repair, or in stem cells activated in the post-developmental period for the first time. Both types of events are hypothesised to contribute to the ageing process.

The view of ageing as the progressive accumulation of pseudo-random events provides a new interpretation for diseases typically associated to the old age (Alzheimer disease, type II diabetes, etc.), which seem to be caused by the malfunctioning of specific genes, but whose onset in humans typically occurs from the 5th decade of life onwards. Our hypothesis is that the difference between these diseases and the effects of “normal”, “healthy”, ageing is more quantitative than qualitative. We think that both ageing and ageing-associated diseases are driven by biological change events caused by genes set to be activated in the old age: the phenotypic manifestations of normal ageing are only milder, more benign than those associated to such diseases. This interpretation explains why the temporal patterns of ageing and
age-related diseases are coincident, and finds a strong confirmation in the fact that measures
known to delay ageing, such as caloric restriction, also postpone the onset of diseases [3].
As we mentioned, a slow-down in the pace of ageing can be obtained through dietary re-
striction. One possible explanation of this evidence is that dietary restriction directly affects
the functioning of a biological equivalent of the global clock, which becomes slower. As a re-
result, all genes bound to be activated in the ageing period are delayed, all by the same amount.
The genetic pathways known to influence life span may be part of the molecular device which
transduces the clock value (perhaps implemented as a diffusible molecule) from the external
acellular environment to the nucleus of cells.

5 Cancer

Cancer is a class of diseases involving unregulated cell growth. According to the “standard
theory”, also referred to as the “multi-hit” hypothesis [20], carcinogenesis is a multi-step pro-
cess that can take place in any cell, driven by multiple damages (“hits”) to genes that normally
regulate cell proliferation and cell death. An alternative theory considers cancer stem cells
responsible for tumour growth [1]. Cancer stem cells share with normal stem cells the ability
to self-renew and differentiate into multiple cell types: the presence in tumours of cells of
different types and/or having different degrees of differentiation is a well documented phe-
omenon, coherent with the cancer stem cell theory.

In recent years, new cancer stem cell models have been proposed, e.g., the “complex sys-
tem model” [4], or the “stemness phenotype model” [21]. These models share the ideas that all
cells in a tumour are potentially tumorigenic, and that stemness is a dynamic property. The
stem-non stem conversion would occur in both directions, triggered by genetic and epigenetic
factors, and influenced by the cellular microenvironment.

A critical point for all cancer theories is the impossibility to reconduct tumour formation
to a common set of gene mutations. A few cancer-related genes, such as p53, do seem to be
mutated in the majority of tumours, but many other cancer genes are changed in only a small
fraction of cancer types, a minority of patients, or a subset of cells within a tumour. Although
the effort to reconduct tumour formation to subsets of mutated genes has been unsuccessful,
it is nonetheless undeniable that correlations between tumours and patterns of mutations
exist, i.e. individual genes are mutated in percentages that are tumour-specific [29].

In ET, as we have seen, some stem cells are activated during development. After activa-
tion, they persist in the body in a dormant condition, and can be reactivated after develop-
ment, in what we called the ageing period, to regenerate damaged body parts. Other stem
cells are activated in the post-developmental period for the first time. The stage for a danger-
ous scenario is set if a fault arises in one of these stem cells, affecting one or more morphogen
processing modules. Let us assume that in a stem cell, which we will call mother cell, the
mobile code-regulatory set match results in a proliferation event. Some of the normal cells
generated are later turned into new stem cells, which we may call daughter cells. The mo-
bile code generation device used by the daughter cells is inherited from their mother: if it is
damaged, the damage will be present also in the daughters.
The damage to morphogen processing modules may impede the creation of the new mobile code element which is responsible for differentiation (Fig. 13): therefore one or more daughter stem cells may end up having the same mobile code as their mother (Fig. 14). Since these stem cells also share the genome with the mother, the same mobile code-regulatory set match that triggered a proliferation in the mother is bound to occur also in the daughters, giving rise to identical proliferations, from which granddaughter cells will be produced, and so on. The result of this scenario is an endless chain of proliferation events, the mark of carcinogenesis. The engine of the process is a set of cancer stem cells, which are dynamically and continuously created, and all contain the initial damage.

In ET, the mobile code generation device is composed of all morphogen processing modules. Therefore, a damage to this mechanism may correspond in biology to mutations affecting genes involved in morphogen processing pathways. Interestingly, genes belonging to such pathways are frequently mutated or aberrantly activated in cancer [17]. In our model, the set of morphogens necessary to induce differentiation are specific for each part of the structure. If we translate this to biology, the same may be true for biological morphogens, and biological tissues and organs.

Given, for instance, four biological morphogens: BMP, Nodal, Notch and Wnt (indicated with the letters a, b, c, d) a subset of the morphogens (say a and b) may be needed for skin differentiation, and another subset (say b, c and d) may be required for gut differentiation. If a sufficient number of morphogen processing genes in one or more pathways are rendered non-functional (through mutations), differentiation fails, and cancer ensues.

Based on studies of skin cancer, the process of carcinogenesis is traditionally divided into three phases: initiation (induction of permanent alterations to DNA), promotion (proliferation of the initiated cell triggered by subsequent stimuli) and progression (from a benign tumour to a malignant one). Our interpretation is that the damage to morphogen processing pathways corresponds to initiation, the first stage of carcinogenesis. The analogous in our model of the promotion phase, i.e. the proliferation of the stem cell, is postponed until regeneration occurs, or until the clock reaches the timer value. The action of a timer-dependent mechanism, if present in biology, may explain the long latency period observed between the exposure to carcinogenic substances (e.g. tobacco smoke) and the appearance of a tumour (e.g. lung cancer). Finally, progression, the third and last phase of carcinogenesis, may be linked to further mutations which confer additional powers to the already transformed cells, i.e. the capacity to infiltrate tissues and to produce distant metastases.

Our proposal is fully consistent with the cancer stem cell model, with the recent paradigm of dynamic stemness, and with the idea that each cell in a tumour has tumorigenic potential. This potential is associated to a damage to morphogen processing genes, which is present initially in a single stem cell, and subsequently spreads to all normal and stem cells derived from it. The ET contribution to the field mainly lies in two ideas. The first idea consists in the notion of stem cell generation, which explains the emergence of a hierarchy of cell types, and brings together embryonic, adult and cancer stem cells under a unified framework. The second idea is a proposal for a low-level generation mechanism, through the actions of
Figure 13: Carcinogenic normal-stem transition in ET. A damage inside the mobile code generation device (composed of morphogen processing modules) impairs the differentiation process. As a result, the mobile code of the daughter cell is the same as the normal cell from which the stem derives, which in turn is the same as the (mother) stem cell from which the normal cell was created.

Figure 14: Cancer stem cells drive tumorigenesis in ET. The Figure reports the tumorigenic version of the development shown in Fig. 7. Thanks to a fault in the mobile code generation device of the mother, one or more daughter stem cells have the same mobile code as the mother (A). Since this mobile code triggers a proliferation for the mother (panel 1A), the same holds true for the daughters (panel 2A), and for the granddaughters, etc., leading to an endless chain of proliferation events. As a result the population of blue cells tends to increase over time. Other cell types (pink and purple), which were generated also during normal development, are still present.
morphogens and the division of space into regions, which may correspond to biological stem niches.

| cancer type | bladder | brain | breast | colon |
|-------------|---------|-------|--------|-------|
| pathways involved | a,b,c | a,b,d | a,c,d | b,c,d |
| combinations of damaged pathways | \{a\} | \{a\} | \{a\} | \{b\} \\
| | \{b\} | \{b\} | \{c\} | \{c\} \\
| | \{c\} | \{d\} | \{d\} | \{d\} \\
| | \{a,b\} | \{a,b\} | \{a,c\} | \{b,c\} \\
| | \{a,c\} | \{a,d\} | \{a,d\} | \{b,d\} \\
| | \{b,c\} | \{b,d\} | \{c,d\} | \{c,d\} \\
| | \{a,b,c\} | \{a,b,d\} | \{a,c,d\} | \{b,c,d\} |

Table 1: Hypothetical involvement of four morphogen processing pathways in as many cancer types.

In our interpretation, morphogens and morphogen processing pathways are organ-specific, and a fully working differentiation machinery requires the functioning of all morphogen processing pathways involved in the generation of the organ. On the other hand, a reduced form of differentiation can be obtained in many different ways, corresponding to all possible different subsets of non-functional morphogen processing pathways. Table 1 gives an overview of a hypothetical involvement of four morphogen processing pathways (a, b, c, d) in four organs / cancer types (bladder, brain, breast, colon). For each organ three morphogen processing pathways are hypothesised to be necessary for differentiation. The row labelled “combinations of damaged pathways” lists all possible combinations of damaged pathways which lead to failed differentiation (and hence tumour formation) in the relevant organ. Different combinations of damaged pathways may correspond to tumours of different grades. Considering that a damage in each morphogen processing pathway can be caused by mutations to a number of genes, this scheme can help to explain the heterogeneity of the mutational patterns observed, and the presence in tumours of cells having different degrees of differentiation and/or of different cell types.

6 Conclusions

We wish to conclude this summary indicating some experiments that could be used to test ET. One of the pillars of the model is the dynamic nature of the normal-stem transition. Stem cells can generate normal cells, and normal cells can revert to stem cells, with a new mobile code. This provides a conceptual bridge that links together embryonic, adult, ageing and cancer stem cells, and explains the emergence of a hierarchy of stem cells, a fact corroborated by data on numerous organs or biological subsystems. In order to validate the model, it is necessary to reproduce the biochemical mechanism which enables the normal-stem conversion, and shed light on the interplay between morphogen processing pathways and HOX genes, which is deemed responsible for mobile code generation.

As we have seen in section 3, germline penetration implements a flow of genetic information from somatic to germline cells, to be passed on to future generations: as such, it can be
considered the carrier of a transposon-mediated inheritance device. As a first step to prove
the existence of the flow, we need to identify the moment in which transposable elements
exit their host stem cells to reach the circulatory system, and make their way towards the
germline. Furthermore, our model predicts that the spread of transposons is an event which
immediately (in evolutionary terms) follows a lineage change. This could be done through
techniques for estimating the age of DNA sequences more sensitive than those currently in
use.

In section 5, we have proposed an explanation for the heterogeneity of patterns of gene mu-
tations observed in different tumours. The explanation is based on a hypothetical differential
involvement of morphogen processing pathways in the generation (and regeneration) of differ-
ent organs. This prediction is susceptible of verification with bioinformatical techniques. To
achieve this objective, all cancer-related genes should be re-classified based on their involve-
ment in morphogen processing pathways. Then, existing databases should be screened for
mutations to genes belonging to each morphogen processing pathway. Finally, data should be
examined, to see if they can support a statistical correlation between number and/or severity
of functional impairment in morphogen processing pathways and tumour grade.

In conclusion, we have shown that ET is able to generate arbitrary 3-dimensional cellu-
lar structures of unprecedented complexity, and can reproduce a simplified version of some
key biological phenomena such as development, the presence of transposable elements in the
genome, ageing and cancer. For this reason, we think it can be proposed as a general-purpose
model for multicellular biology. Future work will be aimed at further reducing the biological
gap, mapping the model variables to individual genes and chemical elements.

References

[1] D. Bonnet and J.E. Dick. Human acute myeloid leukemia is organized as a hierarchy that
originates from a primitive hematopoietic cell. Nature Medicine, 3(7):730–737, 1997.
[2] L. Collier and D. Largaespada. Transposable elements and the dynamic somatic genome.
Genome Biology, 8(Suppl 1):S5, 2007.
[3] R.J. Colman et al. Caloric restriction delays disease onset and mortality in rhesus mon-
keys. Science, 325(5937):201–204, 2009.
[4] M.H. Cruz et al. The stemness phenotype model. ISRN Oncology, 2012.
[5] V.D. Dilman and W. Dean. The neuroendocrine theory of aging and degenerative disease.
Center for Bio-Gerontology, 1992.
[6] A. Fontana. Epigenetic tracking, a method to generate arbitrary shapes by using evo-
devo techniques. In Proceedings of the 8th international conference on epigenetic robotics
(EPIROB), 2008.
[7] A. Fontana. Epigenetic tracking: biological implications. In Proceedings of the 10th
European conference on artificial life (ECAL), volume 5777 of LNCS, pages 10–17, 2009.
[8] A. Fontana. An artificial life model for carcinogenesis. In Proceedings of the 12th in-
ternational conference on the simulation and synthesis of living systems (ALIFE), pages
101–108, 2010.
[9] A. Fontana. Devo co-evolution of shape and metabolism for an artificial organ. In Proceedings of the 12th international conference on the simulation and synthesis of living systems (ALIFE), pages 15–23, 2010.

[10] A. Fontana. A hypothesis on the role of transposons. Biosystems, 101:187–193, 2010.

[11] A. Fontana and B. Wróbel. A model of evolution of development based on germline penetration of new “no-junk” dna. Genes, 3:492–504, 2012.

[12] A. Fontana and B. Wróbel. An artificial lizard regrows its tail (and more): regeneration of 3-dimensional structures with hundreds of thousands of artificial cells. In Proceedings of the 12th european conference on artificial life (ECAL), pages 144–150, 2013.

[13] A. Fontana and B. Wróbel. Embryogenesis, morphogens and cancer stem cells: putting the puzzle together. Medical Hypotheses, 81(4):643–649, 2013.

[14] A. Fontana and B. Wróbel. Pseudorandomness of gene expression: a new evo-devo theory of ageing. Current aging science, 7:48–53, 2014.

[15] R. Gerschman et al. Oxygen poisoning and x-irradiation: a mechanism in common. Science, 119:623–626, 1954.

[16] B.L.M. Hogan. Morphogenesis. Cell, 96(2):225–233, 1999.

[17] C. Karamboulas and L. Ailles. Developmental signaling pathways in cancer stem cells of solid tumors. Biochimica et Biophysica Acta, 1830(2):2481–2495, 2013.

[18] C.J. Kenyon. The genetics of ageing. Nature, 464(7288):504–512, 2010.

[19] T.B. Kirkwood. Evolution of ageing. Evolution, 270 (5635):301–304, 1977.

[20] A. Knudson. Mutation and cancer: statistical study of retinoblastoma. PNAS, 68(4):820–823, 1971.

[21] D.R. Laks et al. Brain tumor stem cells as therapeutic targets in models of glioma. Yonsei Medical Journal, 51(5):633–640, 2010.

[22] L. Liu and T.A. Rando. Manifestations and mechanisms of stem cell aging. The journal of cell biology, 193(2):257–266, 2011.

[23] E.J. Masoro and S.N. Austad. Comparative genetics of aging. Academic Press, 2007.

[24] B. McClintock. The origin and behavior of mutable loci in maize. PNAS, 36(6):344–355, 1950.

[25] K.R. Oliver and W.K. Greene. Transposable elements: powerful facilitators of evolution. Bioessays, (7):703–714, 2010.

[26] J.C. Pearson et al. Modulating hox gene functions during animal body patterning. Nature Reviews Genetics, 6(12):893–904, 2005.

[27] T. Reya et al. Stem cells, cancer, and cancer stem cells. Nature, 414(6859):105–111, 2001.

[28] X. Song et al. Bmp signals from niche cells directly repress transcription of a differentiation-promoting gene, bag of marbles, in germline stem cells in the drosophila ovary. Development, 131:1353–1364, 2004.

[29] C.H. Yeang et al. Combinatorial patterns of somatic gene mutations in cancer. FASEB Journal, 22(8):2605–2622, 2008.