Cell code: the evolving arrangement of molecules that perpetuates life

Antony M Jose

Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD - 20742. e-mail: amjose@umd.edu

Abstract

Our tremendous progress in understanding living things can mask our ignorance of the information needed to perpetuate life. As the basic unit of life, cells hold information in two distinct forms: in the stable DNA sequence that is faithfully replicated during cell divisions, and in the changing arrangement of molecules that return to a similar configuration at the start of each generation. These two interdependent stores of information coevolve while perpetuating an organism and together define the cell code of that organism. Although we do not yet know the entire cell code of any organism, recent developments suggest approaches to discover this code and the mechanisms by which it is reproduced in each generation. Clearly seeing the relationship between an organism and its cell code has implications for the origin and management of disease, evolution, and engineering life.

Introduction. Living things are staggeringly complex and have enchanted generations of biologists, leading to the many approaches that we use to interrogate life processes today. Despite having learned much about living things, we do not know what information is needed to specify the perpetuation of a particular organism. In fact, we have hardly begun to ask the question.

The information needed to perpetuate an organism must be present even in the life stage that has the least number of cells. This stage serves as a bottleneck for the transmission of information from one generation to the next and is minimally a single cell. All the molecules and their arrangement that is reproduced in every such bottleneck stage of an organism is the minimal information required to specify the perpetuation of that organism. I define this information that is nearly reproduced in each generation of an organism as its cell code – a three-dimensional arrangement of molecules present along with the linear sequence information in DNA. Every cell that is capable of generating the entire organism must either have, or be able to reconfigure its contents to generate, the cell code of that organism. Deciphering a cell code requires studies that compare the bottleneck stage of successive generations (e.g. zygote in sexually reproducing organisms). To discover the mechanisms by which a cell code is reproduced in each generation, the entire life cycle of the organism needs to be examined (Figure 1). For chicken, this would mean examining from egg to egg and not from egg to hen (developmental biology) or from hen to egg (reproductive biology). Approaches that enable observation of the bottleneck stage in successive generations and that selectively perturb non-genetic components of a cell will be particularly useful for the discovery and analysis of cell codes.

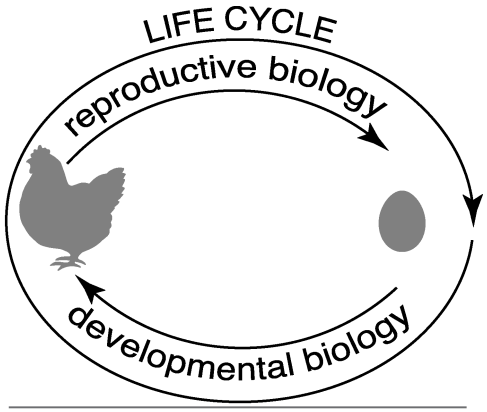

Figure 1. What perpetuates life? To answer this question, the study of how organisms are made (developmental biology) and how they reproduce (reproductive biology) are inadequate. We need to study the entire life cycle.

While this article has an emphasis on animal biology to give it focus, the concepts presented here are applicable in general to all cellular life: bacteria, archaea, and eukaryotes.

Organisms cycle minimally through a single cell. The continuity of life relies on cycles. Single-celled organisms go through cell division cycles (Murray and Hunt, 1993). Parasites go through elaborate cycles within different cells and even different host organisms (Olsen, 1986). Multicellular organisms cycle through a single-cell bottleneck – the fertilized zygote in sexually reproducing organisms – that carries all the information necessary to make the next generation (Wilson, 1896). In animals the zygote can then either be present within an egg that is laid or be implanted within the uterus of the mother and grow as a fetus. In animals that lay eggs, all the information necessary for development is contained within the egg. Under permissive environmental conditions, a hatchling is inevitable. In animals that give birth to live young, the growing fetus shares circulation with the mother, which can be thought of as providing the permissive environment and more. Organisms that undergo parthenogenesis (Phillips, 1903) also cycle through a single cell, an egg that does not need fertilization by sperm. Finally, regenerating organisms (Birnbaum and Sanchez-Alvarado, 2008) can begin each generation from many different collections of cells. Nevertheless, in every scenario, the simplest life stage of an organism minimally consists of a single cell.

A single-cell bottleneck constrains multicellularity. The 'beginning' of an organism's life is an arbitrary time-point in its life cycle (McLaren, 1980). For multicellular organisms that

reproduce sexually, if we consider the union of gametes to generate the single-cell zygote to be the beginning, what organisms can a zygote develop into while ensuring the continuity of life? An initial answer could be anything that is compatible with the production of the zygote for the next generation. That is, any collection of cells can develop as long as they do not interfere with the ability to reproduce. Indeed, multicellular organisms have evolved many different types of cells with very different properties (Arendt et al., 2016) – consider the sciatic nerve that can grow to be a meter long versus a tiny lymphocyte in humans. In most cases the DNA within every cell is essentially the same. The drastic differences between cells are achieved by changing how the same DNA is used to make RNA, protein, and other components within a cell. As a powerful demonstration of this principle, the nucleus of any cell can be combined with the cytoplasm of an oocyte to recreate an entire animal (Briggs and King, 1952; Gurdon et al., 1958) and the addition of a few factors can reprogram one cell type into another (Lassar et al., 1986; Takahashi and Yamanaka, 2006). Notable exceptions to this general rule include human red blood cells, which jettison their nuclei as they mature (Thompson, 1951). Nevertheless, the collection of remarkably different cells that can make up an organism poses no problems for the continuity of life as long as it is compatible with the creation of a zygote to start the next generation, hence the famous saying "a hen is only an egg's way of making another egg" (Butler, 1910). Expanding these considerations to populations, however, reveals that individual members of a population can be exempt from this constraint as long as they support one individual that is bound by the constraint to produce a zygote. For example, in colonies of bees or ants, drones or workers can be sterile and yet support the queen, which perpetuates the colony (Heinze and Schrempf, 2008). These constraints focus our attention on the cell – the zygote in many cases – as an important unit of organization that needs to be understood to explain life.

Much of the information needed to specify the zygote of any organism is unknown. Factors that identify a specific zygote include both the contents of the cell and the spatial arrangement of the contents, which can be crucial for function. Imagine a "typical" eukaryotic cell with many typical components like mitochondria, lysosomes, endoplasmic reticulum, and other organelles; ribosomes, polymerases, and other molecular machines; ATP, Ca^{2+}, and other small molecules and so on. Everything in this cell needs to be organized in three dimensions for appropriate function. *What would we need to add to our description of this typical cell to describe the zygote that can develop into a particular organism?* The answer to this question is likely to include many different molecules. For example, a genome with appropriate chromatin to allow access to early transcription factors, mRNAs that are poised to be translated, and sufficient stores of small molecules and enzymes that use the molecules to initiate development in permissive environments. Each of these molecules likely needs to be present at precise amounts as illustrated by the control of mRNA levels in the zygote, which is tuned to allow development (Gerson-Gurwitz et al., 2016). Attempts have been made to comprehensively catalogue different components of cells: the genome, the epigenome, the transcriptome, the proteome, the metabolome, etc. These techniques can be used to discover the contents of a particular zygote. Discovering the spatial arrangement of components within a cell, however, has been more difficult but attempts to generate a comprehensive map of at least the proteins within a few cell types are underway (Thul et al., 2017; Allen Cell Explorer). Yet, spatial arrangement and sub-cellular localization are crucial for function. Furthermore, we now know that the genome is organized into spatial domains that impact gene expression in eukaryotes (reviewed in Bonev and Cavalli, 2016). The functions of key molecules within cells are regulated through chemical modifications that continue to be discovered (e.g. RNA modifications (Helm and Motorin, 2017), DNA modifications (Sood et al., 2016), histone modifications (Andrews et al., 2016), etc.). Finally, we currently do not understand how the components within a cell determine many aspects of cellular behavior (see Marshall, 2015 for an interesting list). Thus,

there is much to learn before we can specify the molecules and their arrangement that describes the zygote of any organism.

DNA proposes, cell disposes. While the genome is a repository of all sequences that can be transcribed and used for making other components, at any given moment, which RNA is transcribed from DNA depends on what else is within the cell and the cell's interactions with the external environment. The information contained in the genome of an organism is thus not sufficient to make that organism. To appreciate this insufficiency, consider a single cell in an organism: (a) the DNA within this cell does not encode all aspects of all molecules in the cell; and (b) whether a molecule is made using the DNA depends on other contents in the cell.

The linear DNA sequence is not sufficient to build the three dimensional cell: The RNA and protein components of a cell ultimately rely on DNA for generation through the process of transcription and translation using smaller molecules, nucleotides and amino acids, as raw materials. These proteins and RNA can go on to catalyze the formation of additional components such as lipids and sugars from other raw materials. However, the three-dimensional structure and spatial arrangement of all the molecules within a cell is not encoded in the linear DNA. For example, where a protein is localized in the cell changes over time based on interactions with other components within a cell (Thanbichler and Shapiro, 2008). Components in organelles such as mitochondria are not entirely encoded by the nuclear genome and can be considered as endosymbionts with partial autonomy within a eukaryotic cell (Mereschkowsky, 1905; Sagan, 1967). While the DNA encodes the sequence of a protein, the structure it takes depends on its immediate environment (Englander and Mayne, 2014). Finally, sizes, numbers, and shapes of organelles, and indeed the collective three-dimensional architecture of a cell all depend on dynamic interactions between components within the cell and with its external environment independent of the DNA (Rafelski and Marshall, 2008).

Many cell types use the same genome: The presence of different cell types that retain their identity over time and across cell divisions within an organism reveals that the contents of a cell can exert a controlling effect on the DNA. The mechanisms that make and maintain different cell types were initially referred to as epigenetic control systems (Nanney, 1958) to signify that they are above ('epi') genetic control. Pioneers studying such control mechanisms in bacteria imagined six modes of regulation for components within a cell in the context of biochemical reactions (Monod and Jacob, 1961): (i) cross feedback, (ii) feed forward loop, (iii) mutual inhibition, (iv) mutual stimulation, (v) signal-dependent reversible activation, and (vi) oscillation. These regulatory modes provide teleonomic (i.e programmed) constraints that explain "how come?" rather than "what for?" when events happen within a cell (Mayr, 1961). Numerous examples for all of these imagined modes of regulation have been identified in the past half-century. Furthermore, computational exploration suggests that even random networks of elements with high molecular specificity can result in the emergence of different cell states that remain stable over time (Kauffman, 1969). Extensions of these ideas have led to the identification of gene regulatory networks in model organisms (Levine and Davidson, 2005), where collections of active genes and gene products maintain different cell types while being able to respond to signals during development.

The cell code. Since its beginnings as *Entwickelungsmechanik* more than a century ago (Sander, 1991), the rich field of developmental biology has been addressing various aspects of how an organism is made. But, the minimal information that is necessary to specify the development of a particular organism in a given external environment is still unknown, and is in fact, relatively unexplored.

Consider an organism that progresses from one generation to the next through a single-cell stage (Figure 2A), as is the case with humans and most known multicellular organisms. Let

us define everything that is produced using the DNA within the unicellular stage – macromolecules plus organelles plus spatial arrangement – as "C" and the interacting aspects of the environment as "E" (Figure 2B). These external inputs include factors that influence (e.g. temperature), interact with (e.g. extracellular signal), or are taken up (e.g. sugars) by the cell. At an arbitrary initial time, t = 0, C_0 includes RNA and proteins made using the DNA and other components present in the cell at that time and E_0 includes molecules in the external environment (which includes other cells) that is independent of the DNA within the cell. At this arbitrary initial stage these components can interact and/or modify each other (depicted as •). As time progresses, everything in the cell is free to change while the DNA is being used to make new RNA and proteins, and more external material can interact with the cell or be imported into the cell. That is, at time t = 1, C_0 made at t = 0 has changed ($C_{0\rightarrow 1}$) and can now interact with and/or modify (•) C_1 made at t = 1. Similar changes can also occur in the imported or interacting external material ($E_{0\rightarrow 1}$• E_1). Furthermore, both the cellular output and the external input can interact and/or modify each other. These changes over time include production of new material (e.g. protein synthesis, lipid synthesis, etc.) and destruction or modification of old material (e.g. autophagy, phosphorylation of lipids, etc.). This process of iterative accumulation of material and progressive change in the state of a cell implies that the DNA *and* everything else accrued in the past that is now within the cell together predicts the next state of a cell.

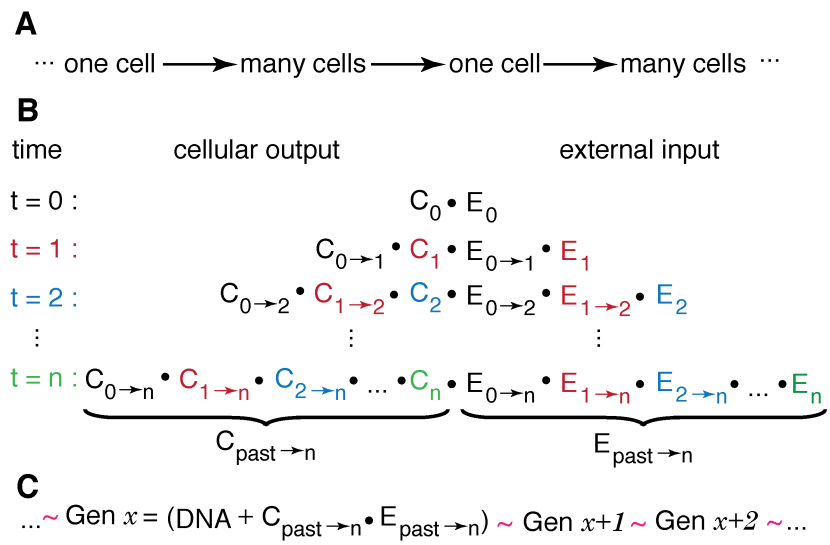

Figure 2. Perpetuation of life. (**A**) Organisms are minimally unicellular and most multicellular organisms cycle through a single-cell bottleneck. (**B**) Every cell accrues components guided by the action of preexisting components on DNA (C_0, C_1, C_2, ...). The new components made at each moment (t) can interact (•) with preexisting components, which could have changed since they were made (e.g. $C_{0\text{->}2}$ and $C_{1\text{->}2}$ at t = 2). This process extends to external components the cell interacts with or acquires independent of DNA within the cell (E_0, E_1, E_2, ...). At a given time (t = n), the accrued cellular output ($C_{past\text{->}n}$) interacts with the accrued external input ($E_{past\text{->}n}$). (**C**) DNA and a specific arrangement of components accrued over time that are together required to make an organism and contained within the single cell that begins each generation (Gen x) defines *the cell code* of that organism. Recreation of a similar (~) code in successive generations perpetuates life and is a hallmark of organisms that constrains their evolution.

Early attempts at unifying biology focused on understanding the cell in development and inheritance (Wilson, 1896). More than a century ago, some experimental embryologists clearly appreciated the iterative progression (e.g. Boveri, 1902) and influence of history (e.g. Boveri, 1906) when describing the nature of organisms. The interdependence of genes and non-genetic factors within a cell was also well appreciated (e.g., Conklin, 1908; Waddington, 1962). In fact, the term 'epigenetics' was initially defined as "…causal mechanisms by which the genes of the genotype bring about phenotypic effects" (Waddington, 1942), but later stated as "the causal interactions between genes and their products which bring the phenotype into being" (Waddington, 1968), making it clear that the genotype (DNA) alone does not bring the

phenotype into being. These insights from embryology were also recognized by early proponents of the Modern Synthesis that combined Mendelian genetics with Darwinian evolution (Huxley, 2010), and when ignored, can lead to the mistaken popular view that DNA is the blueprint of life.

In the context of an organism, the contents of a cell can change dramatically over time independent of the DNA within the cell. For example, consider progression from one generation to the next in the worm *C. elegans* (Hubbard and Greenstein, 2005) – the best characterized animal we know. From the single-celled zygote, two primordial germ cells are established five cell divisions later. These two cells then go through an extended period of quiescence while the rest of the organism develops. Towards the end of this period intestinal cells cannibalize a large portion of the cytoplasmic contents of both cells (Abdu et al., 2016). Then, the cells proliferate first producing sperm and then oocytes. Maturing oocytes acquire cytoplasmic contents from the rest of the germline (Wolke et al., 2007) and yolk from the intestine (Grant and Hirsh, 1999) potentially along with extracellular RNAs (Marré et al., 2016). These acquisitions make the oocyte larger than all the 558 cells of a hatching larva combined. Fertilization of this enlarged oocyte creates the zygote for the next generation. Both dramatic changes in this cycle – loss of cytoplasmic material from primordial germ cells and gain of cytoplasmic material by oocytes – occur during periods of relative transcriptional quiescence.

Explicitly considering the influence of time and loss/acquisition of non-genetic material, we can depict the single cell that begins each generation at an arbitrary time t = n as DNA plus components accrued until that time and their three-dimensional arrangement i.e. DNA plus $C_{past \rightarrow n} \bullet E_{past \rightarrow n}$. Everything in this cell that is nearly reproduced in the zygote of successive generations is the cell code for making the organism. The upper limit for the cell code is the entire contents and their arrangement in the zygote that begins each generation. In other words, maximally the cell code of the organism = [DNA + $C_{past \rightarrow n} \bullet E_{past \rightarrow n}$] (Figure 2C). However, the cell code need not equal everything in a zygote. If only a subset of the components and arrangement within a zygote are reproduced in every generation (Figure 3), then the cell code can be less than the contents of a zygote. Furthermore, because many molecular assemblies in life are capable of self-organization (e.g. mitotic spindle (Salmon, 1975)) and templated processes (e.g. transcription (Weiss and Gladstone, 1959)), it is possible that essentially the same cell code could be specified using a subset of the molecules and arrangement that are recreated in every zygote. This possibility is supported by experiments on single-cell regeneration in *Stentor polymorphus* - 1/27th of its initial volume can regenerate all structural features (Lillie et al., 1896).

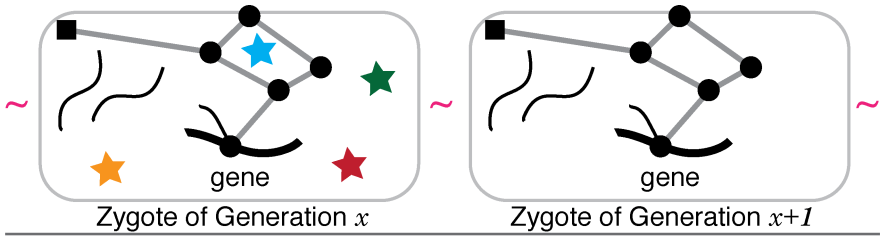

Figure 3. Nature of the cell code. Part of the cell code from the perspective of a single gene (black) and its RNA (curved lines) and protein (square) is depicted. Complexes of other proteins/RNAs/lipids etc. (circles) potentially localized in specific places within the cell and interactions between them (grey lines) that can impact the gene and that are reproduced in every generation are part of the cell code. Some components in the zygote may not be made in successive generations (stars) and compensatory changes in the rest of the zygote may permit different configurations for this gene while preserving the overall cell code of the organism.

Determining the minimal representation of the cell code of an organism requires asking and answering a set of questions at the molecular level that have rarely, if ever, been posed in experimental systems. For example, imagine a set of details that we have determined about a

gene and its products in an organism that are all reproduced in the zygote of every generation, i.e., that are all part of the cell code. Lets say this gene has chromatin modifications, 2 copies of partially transcribed RNA, 32 copies of the spliced RNA, 17 copies of translated protein, 12 copies of the protein with correct subcellular localization etc. Which of these details are necessary for the perpetuation of a similar arrangement? What changes to the rest of the genome can impact this gene and the arrangement of its products while perpetuating life? What other arrangements of these components are essentially equivalent?

A problem that is separate from determining the cell code of an organism is determining how it is reproduced in every generation. Reproduction of the genetic component of the cell code, i.e. DNA, is trivial and is simply through the process of DNA replication at each cell division. Reproduction of the non-genetic components of the cell code, i.e. the spatial arrangement of molecules including DNA, is likely to be more complex. One possibility is that these arrangements are preserved through cell divisions despite differentiation and development of the organism such that a continuous line of cells from the zygote of one generation to the next carries the cell code for the organism. Alternatively, these arrangements could vary during differentiation and development but go through an elaborate cycle such that they return to a similar configuration in the zygote of each generation. In the case of unicellular organisms, the cell code needs to be reproduced at a certain time in each cell division cycle. In organisms such as plants where many cells are capable of generating the entire organism, every such cell needs to be able to reproduce the cell code when exposed to conditions that stimulate differentiation and development. Currently, we simply do not know the mechanism by which the cell code is reproduced in any organism.

A range of components and arrangements are likely to be nearly equivalent cell codes for an organism. We already know that the genomes of individuals within a species can vary a little and yet result in the generation of similar organisms. Discovering other alterations to the cell code that are compatible with the perpetuation of life will reveal the full extent of novelty that can arise from one generation to the next in an organism.

Persistent ancestral information modifies the cell code. A key concern in thinking about how organisms evolve is what changes in one generation can be passed on to subsequent generations. The nature of organisms presented here (Figure 2 and Figure 3) makes it clear that the persistence of changes across many generations requires modification of the cell code and thus evolution occurs through “descent with modification” (Darwin, 1859) of the cell code. Note that this does not include information (induction by signals, uptake of nucleic acids, etc.) accrued from additional sources such as the microbiome or maternal circulation during development that could also vary across generations.

Modifications to the cell code can alter its genetic (DNA) or non-genetic aspects (molecules and/or arrangement). Changes to the DNA sequence can persist for many generations and are indeed the best-studied changes to the cell code. These can include both random mutations and changes derived from experience (e.g. CRISPR-Cas system of antiviral immunity (Barrangou, 2015)). But, changes to the cell code that do not alter DNA sequence also have the potential to persist for many generations. In fact, one of the earliest mutants ever described - in 1744 - was found to not alter the DNA sequence. This variant of the toadflax *Linaria vulgaris* called *peloria* (greek πελωρ, monster) (Linnaeus and Rudberg, 1744; Gustafsson, 1979) arises at ~1% frequency in each generation (DeVries, 1906) and is correlated with methylation but not sequence changes at the *Lcyc* locus (Cubas et al., 1999). Early work in the ciliate *Paramecium aurelia* clearly demonstrated that changes in the cortex created using explants could be transmitted for many generations (Beisson and Sonneborn, 1965). Some changes reverted after 30-40 cell generations, having successfully transmitted the changed organization of the cortex to 10^9 to 10^{12} cells, and other changes could apparently be

maintained indefinitely. Because many ciliates develop using RNA from one generation to splice or unscramble germline DNA sequences in the next generation, changes in RNA can alter how the DNA is rearranged in many subsequent generations (Bracht et al., 2013; Chalker et al., 2013). In other organisms, extensively studied phenomena like paramutation (Brink, 1956; reviewed in Hollick, 2017) and RNA interference (Fire et al., 1998) that cause persistent silencing of genes also illustrate the persistence of non-genetic information. While the precise mechanisms for how the gene silencing information is transmitted across generations are still being worked out, persistent chromatin modifications and amplified RNAs have emerged as possibilities (reviewed in Rankin, 2015). Transmission for many generations is likely to require inheritance of molecules containing the information for silencing a gene for one or a few generations combined with periodic replication of molecules to reinforce this information. For example, an ancestral event such as exposure to double-stranded RNA during RNA interference could trigger production of chromatin modifications in each generation based on instructions held in RNA. All that is needed to make this information stable for many generations is transmission, even for just one generation, followed by reproduction in every generation. Crucially, the resulting permanent changes to the cell code do not involve mutating the DNA. Other mechanisms for changing the cell code without mutating DNA include the transmission of prions (Halfmann and Lindquist, 2010), where alternative folding states of proteins that can template recreation of similar states are transmitted across generations.

When complex organisms, like humans, that can make tools and artifacts by interacting with both the environment and other organisms are considered, there is no end to the longevity of ancestral information. For example, consider the impact of ideas in books on the lives of people. The spread of such ideas within a culture (i.e. a meme) can have a profound influence on humans (Dawkins, 1989). Thus, the evolution of such organisms is shaped by much more than their cells (Jablonka and Lamb, 2005). In fact, contributions by many biological mechanisms that transmit information across generations can be incorporated into a consideration of how organisms evolve (Rivoire and Leibler, 2014), if unconstrained by evidence.

Even if only the transmission of biological material is considered, the precise way in which information is transmitted across generations is unclear. While DNA in a zygote is obtained only from the immediate parent, the provenance of the information contained in the non-genetic material with a zygote is less clear. If life originated once on earth as suggested by the commonalities among extant organisms, then every zygote can be traced back to the first cell by descent. This initial cell must have accrued complexities that were transmitted from one generation to the next. Therefore, although it is possible that the information contained in the non-genetic components and their arrangement is also from the immediate parent, it is also conceivable that this information initially arose in an ancestral generation and has since been faithfully propagated. Nevertheless, unlike in the case of genetic changes, there have only been a few cases of clear evidence for non-genetic changes that persist for many generations (see Jablonka and Raz, 2009 for review of early work). This paucity could reflect limited experimentation, incompatibility of the changes with the perpetuation of life, or mechanisms that oppose such persistence.

Forces that oppose variation preserve the cell code. The developmental program of an organism imposes constraints on non-genetic variations within the germline, which conveys the cell code from one zygote to the next. The development of an organism in a given environment follows a path that reflects the teleonomic constraints imposed by epigenetic control systems within cells and cell-cell interactions in that organism. This constrained path was referred to as “chreod” (“necessary path” from greek roots χρη, it is necessary, and όδοζ, path (Waddington, 1957)) and has been given modern form within the framework of dynamical systems theory

(Pisco et al., 2016). Such developmental constraints have been recognized as one of the forces that oppose change in organisms during evolution (Gould and Eldredge, 1993).

Some studies show correlation of certain molecular changes in an animal with environmental or dietary changes in parents or recent ancestors (e.g. Rodgers et al., 2013; Rechavi et al., 2014; Sharma et al., 2016; Chen et al., 2016; Klosin et al., 2017) and molecular changes that occur within germ cells can cause effects that persist for a few generations (e.g. Greer et al., 2011; Siklenka et al., 2015). These instances of heritable changes reflect non-genetic modifications of the cell code and urge consideration of the possible ancestral origins of disease and the impact of medical intervention on descendants. However, loss or erasure of such chemical changes after a few generations reflects a homeostatic return to the original cell code. For example, during early mammalian development most parental chemical modifications are erased and new modifications are added (Feng et al., 2010). Because this happens in every generation, the information for adding the new modifications must be passed from one generation to the next as an aspect (molecule and/or arrangement) of the cell code. This developmental reprogramming thus preserves the cell code and opposes transgenerational epigenetic changes. Nevertheless, cases of persistent non-genetic changes - some lasting for tens to hundreds of generations (e.g., Vastenhouw et al., 2006; Ashe et al., 2012; Buckley et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012; de Vanssay et al., 2012; Devanapally et al., 2015; Leopold et al., 2015; Lev et al., 2017; Ciabrelli et al., 2017; Devanapally et al., 2017) - provide us with opportunities to analyze non-genetic aspects of the cell code (Figure 4).

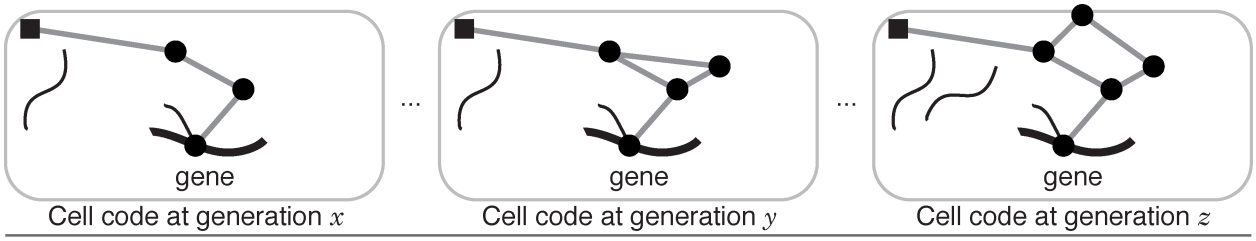

Figure 4. Ancestral non-genetic information and the evolution of the cell code. Schematic depicting possible evolution through changes to non-genetic components of the cell code from the perspective of a single gene. Past non-genetic changes (in generations x, y, z, say) impact the evolving cell code *if* they persist through life cycles *and* overcome homeostatic mechanisms that oppose changes to the cell code.

Together, these considerations inform attempts to identify past events that have shaped the evolution of organisms and future attempts to synthesize organisms.

To make life we need to know the cell code. Considering what we would need to know to make an organism from raw materials reveals the extent of our ignorance of the cell code. While we do not yet know the cell code for any organism, we have begun to remove, replace, augment, or modify aspects of the cell code. In particular, techniques for manipulating DNA have advanced to the point that practically any change in DNA sequence can be made using genome engineering (reviewed in Komor et al., 2017) and even the entire genome of a simple organism can be replaced with a synthetic genome (Gibson et al., 2010). We can also effectively transfer genetic material to avoid mutant mitochondria (Tachibana et al., 2009), we can augment the genetic code (Ostrov et al., 2016), and even use gene drives to change wild populations (Gantz and Bier, 2015). All these efforts are akin to modifying a machine by tinkering with or replacing parts in a pre-existing machine without fully understanding how the machine is put together. As with complex machinery, such manipulations without deep understanding can be dangerous and additionally all manipulations of life warrant careful ethical considerations (e.g. Baltimore et al., 2015). To make a machine from raw materials, however, we need to know the *entire* design. Because organisms build themselves, assembling the cell code of an organism could be sufficient to make an organism. The enormity of this challenge is

clear in our struggle to use raw materials to make a rudimentary cell that perpetuates itself (Blain and Szostak, 2014; Schwille, 2015). Nevertheless, we have begun to coax pre-existing cells into making complex parts of organisms in vitro – e.g. eye (Eiraku et al., 2011) and brain (Lancaster et al., 2013). Eventually making entire complex organisms that reproduce will require first discovering the cell code of a simple organism and then determining the simplest version of that code. The "endless forms most beautiful and most wonderful" (Darwin, 1859) found in nature each have a different cell code. Such organisms sculpted by evolution, however, are unlikely to already have the minimal essential cell code because evolution tinkers with what exists (Jacob, 1977) but anticipates poorly. Our current understanding suggests that evolution proceeds through measure and counter measure including nonadaptive processes (e.g. Lynch, 2007; Madhani, 2013), resulting in a cellular architecture that is more Rococo than Bauhaus with aspects that are superfluous (Gould and Lewontin, 1979). Thus, organisms found in nature have layers of post hoc regulation that can obscure essential design principles. Understanding such principles can enable the design of organisms free of the historically contingent tinkering of evolution.

Approaches to decipher the cell code. Discovering the cell code of an organism minimally requires comparison of features in zygotes of successive generations (Figure 3). Our increasing ability to perform molecular studies on single cells (Tsioris et al., 2014) makes this approach reasonable. Parental effect mutants can be used to identify potential components of the cell code that are maternally or paternally deposited into the zygote. Such mutants may also identify signaling from one generation to the next. However, the arrangement of these molecules in the zygote is more difficult to discover and would require systematic cell biological and biochemical analyses. After these components are identified and their arrangements are discovered, the zygotes of successive generations can be compared.

Penetrating insights often require perturbation of the system and not mere observation. Past perturbation studies of development and reproduction have largely focused on the analysis of essential genes that impact viability or fertility, respectively. Defects in an essential gene (required for development, say) would kill the organism preventing us from examining the zygote in the next generation. Thus, when we interfere with an essential gene or process, we lose the ability to see its impact on the next generation because the intervening organism is affected. In other words, we cannot know if an essential gene or process is required to make the organism or to reproduce the cell code or both. Studies using viable and fertile mutants on the other hand permit examination of the zygote in successive generations. But, careful subsequent analysis will be needed to separate defects in the mere making of the organism from perturbations of the cell code. For example, a mutation in DNA that changes a residue in hemoglobin alters genome sequence in the cell code and affects the structure of the protein made in red blood cells, but likely does not affect non-genetic aspects of the cell code or its propagation. However, a mutation in DNA that changes a protein that is reproducibly present in the zygote or the germline changes not only the genome sequence in the cell code but potentially also non-genetic aspects of the cell code and/or its propagation. Finally, a non-genetic change (e.g. chromatin modification or addition of double-stranded RNA) that causes gene silencing that persists for many generations alters non-genetic aspects of the cell code and potentially its propagation without changing the genome sequence.

The ability to make random changes in DNA and examine its consequences – forward genetics – and more recently to turn off any gene through RNA interference – reverse genetics – were crucial for correlating changes in the genome with changes in the organism. To correlate changes in non-genetic components of the cell code with changes in the organism, additionally we need forward epigenetics and reverse epigenetics. While we do not yet have a way to perform forward epigenetics, reverse epigenetics has begun (Park et al., 2016). For example, a

guide RNA and the Cas9 enzyme fused to a histone modifier can be used to target chromatin modifications to histones located in a specific region of the genome (e.g. Hilton et al., 2015). Such a manipulation could be performed in one generation and its impact analyzed in subsequent generations to determine whether the manipulation altered non-genetic aspects of the cell code.

Nuclear transplantation experiments were essential for the realization that most cells in an organism retain the same genome (Briggs and King, 1952; Gurdon et al., 1958). A similar approach could be used to discover all the cells that retain the ability to use the DNA genome to generate the entire organism, i.e., to discover all the cells that have the non-genetic aspects (molecules and arrangement) of the cell code. Every cell that can be induced to generate the entire organism is expected to have the entire cell code, potentially in a configuration that is different from that found in the bottleneck stage at the start of each generation during normal development. For sexually reproducing organisms, minimally every cell that is continuous within the female germline from one zygote to the next is expected to have all the non-genetic aspects of the cell code except those obtained from the male gamete upon fertilization. Conversely, every cell that is continuous within the male germline from one zygote to the next is expected to have all the non-genetic aspects of the cell code except those obtained from the female gamete upon fertilization. It will be interesting to discover if other somatic cells retain non-genetic aspects of the cell code and how the cell code is reduced from being present in its entirety in the zygote to the portion in each gamete awaiting union upon fertilization.

An alternative to perturbation studies that overcomes the impasse of lethality or sterility when the cell code is disturbed is the use of tracer studies. To illustrate this approach, imagine you wanted to discover the extent of the circulatory system of an organism. A perturbation study could involve making cuts throughout its body and examining if blood spurts out. A tracer study involves injecting an inert molecule that permeates the entire circulatory system and then imaging that molecule. Therefore, taking the tracer approach to examine the cell code of an organism, we could insert benign sequences (e.g. encoding a fluorescent protein) into its genome, add or remove various regulatory features, and examine if these changes have consequences that persist across generations.

Simpler organisms likely have simpler cell codes. Single-celled organisms like the small eukaryote *Ostreococcus tauri* (Derelle et al., 2006) or bacteria could be chosen as model systems to discover the simplest cell codes. Simpler still are organisms that result from efforts to generate bacteria that have a minimal genome (Hutchison et al., 2016; Glass et al., 2017). While the analysis of these organisms would reveal the logic of a cell code for single-celled organisms, it is conceivable that the need for differentiation in complex animals and plants results in fundamentally different strategies for the assembly and reproduction of the cell code.

To make rapid progress in understanding the principles of the cell code, we are likely to benefit from focus. Examples from the history of biology support this assertion: the discovery of the genetic code began with a 'tracer' study where one RNA sequence (poly-U) was used to synthesize one peptide sequence (poly-Phe) in an *in vitro* translation system (Nirenberg and Matthei, 1961) and the basic principles of gene regulation were worked out through sustained effort on a few genes – e.g., the *lac* operon (Jacob and Monod, 1961). Similarly, focusing on one or a few genes, their protein and/or RNA products, and the factors that impact their recreation in successive generations will allow us to ask specific questions and generate a preliminary understanding of the cell code. This focused approach could be an effective complement to comprehensive approaches that compare many components of zygotes in successive generations in response to experimental manipulation.

Conclusion. The sequence of a genome can be used to identify an organism. However, the genome sequence is not sufficient information for making that organism. To make an organism we need to know its cell code, which is the evolving arrangement of molecules that is nearly reproduced in every generation. Building on more than a century of work in biology, we can now begin to decipher the cell code of an organism by analyzing single cells that start successive generations.

Acknowledgements

The ideas presented here were honed through long conversations with my friend Katerina Ragkousi, through lectures to students of a course on Ignorance run by José Feijo, and through a faculty chalk talk to colleagues at the University of Maryland. I thank Steve Wolniak for initially pointing me to relevant books on cell biology; Karen Carleton, David Mosser, Zhongchi Liu, Caren Chang, Norma Andrews, José Feijo, Wenxia Song, Carlos Machado, Tom Kocher, Charles Delwiche, William Snell, Steve Wolniak, and Sougata Roy for comments on the manuscript; all past and present members of the Jose lab for lively discussions; and my friend John Horner for early support when these ideas were in their infancy. I apologize to any colleagues whose work could not be highlighted in this short article.

Funding

Research in the author's lab is supported through funds from the University of Maryland and from the NIH (R01GM111457).

References

Abdu Y, Maniscalco C, Heddleston JM, Chew TL, Nance J. (2016) Developmentally programmed germ cell remodelling by endodermal cell cannibalism. *Nat Cell Biol.* 18(12):1302-1310.

Allen Cell Explorer. http://www.allencell.org

Andrews FH, Strahl BD, Kutateladze TG. (2016) Insights into newly discovered marks and readers of epigenetic information. *Nat Chem Biol.* 12(9):662-8.

Arendt D, Musser JM, Baker CV, Bergman A, Cepko C, Erwin DH, Pavlicev M, Schlosser G, Widder S, Laubichler MD, Wagner GP. (2016) The origin and evolution of cell types. *Nat Rev Genet.* 17(12):744-757.

Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, Cording AC, Doebley AL, Goldstein LD, Lehrbach NJ, Le Pen J, Pintacuda G, Sakaguchi A, Sarkies P, Ahmed S, Miska EA. (2012) piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. *Cell.* 150(1):88-99.

Baltimore D, Berg P, Botchan M, Carroll D, Charo RA, Church G, Corn JE, Daley GQ, Doudna JA, Fenner M, Greely HT, Jinek M, Martin GS, Penhoet E, Puck J, Sternberg SH, Weissman JS, Yamamoto KR. (2015) Biotechnology. A prudent path forward for genomic engineering and germline gene modification. *Science.* 348(6230):36-8.

Barrangou R. (2015) The roles of CRISPR-Cas systems in adaptive immunity and beyond. *Curr Opin Immunol.* 32:36-41.

Beisson J, Sonneborn TM. (1965) Cytoplasmic inheritance of the organization of the cell cortex in paramecium aurelia. *Proc Natl Acad Sci USA.* 53:275-82.

Birnbaum KD, Sanchez-Alvarado A. (2008) Slicing across kingdoms: regeneration in plants and animals. *Cell.* 132: 697-710.

Blain JC, Szostak JW. (2014) Progress toward synthetic cells. *Annu Rev Biochem.* 83:615-40.

Bonev B, Cavalli G. (2016) Organization and function of the 3D genome. *Nat Rev Genet.* 17(11):661-678.

Boveri T. (1902) Über mehrpolige Mitosen als Mittel zur Analzyse des Zellkerns. Verhandlungen der physicalisch-medizinischen Gesselschaft zu Würzburg. Neu Folge 35: 67-90. Translated by Gluecksohn-Waelsch S (1964) for Foundations of Experimental Embryology. B. H. Willer and J. Oppenheimer, eds. Prentiss-Hall, Inc., Englewood Cliffs, NJ.

Boveri T. (1906) Die Organismen als Historche Wesen. (Organisms as historical beings). Kgl. Universitatsdruckerei von H. Sturtz.

Bracht JR, Fang W, Goldman AD, Dolzhenko E, Stein EM, Landweber LF. (2013) Genomes on the edge: programmed genome instability in ciliates. *Cell.* 152(3):406-16.

Briggs R, King TJ. (1952) Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. *Proc. Natl. Acad. Sci. USA.* 38:455–463.

Brink RA. (1956) A Genetic Change Associated with the R Locus in Maize Which Is Directed and Potentially Reversible. *Genetics.* 41(6):872-89.

Buckley BA, Burkhart KB, Gu SG, Spracklin G, Kershner A, Fritz H, Kimble J, Fire A, Kennedy S. (2012) A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature.* 489(7416):447-51.

Butler S. (1910) Life and Habit. A. C. Fifield: London. (originally published in 1878). p. 134.

Chalker DL, Meyer E, Mochizuki K. (2013) Epigenetics of ciliates. *Cold Spring Harb Perspect Biol.* 5(12):a017764.

Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, Feng GH, Peng H, Zhang X, Zhang Y, Qian J, Duan E, Zhai Q, Zhou Q. (2016) Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science.* 351(6271):397-400.

Ciabrelli F, Comoglio F, Fellous S, Bonev B, Ninova M, Szabo Q, Xuéreb A, Klopp C, Aravin A, Paro R, Bantignies F, Cavalli G. (2017) Stable Polycomb-dependent transgenerational inheritance of chromatin states in Drosophila. *Nat Genet.* E-pub. Apr 24.

Conklin EG. (1908) The mechanism of heredity. *Science.* 27(681):89-99.

Cubas P, Vincent C, Coen E. (1999) An epigenetic mutation responsible for natural variation in floral symmetry. *Nature.* 401(6749):157-61.

Darwin, C. 1859. *On the Origin of Species.* First edition. John Murray, London.

Derelle E, Ferraz C, Rombauts S, Rouzé P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynié S, Cooke R, Saeys Y, Wuyts J, Jabbari K, Bowler C, Panaud O, Piégu B, Ball SG, Ral JP, Bouget FY, Piganeau G, De Baets B, Picard A, Delseny M, Demaille J, Van de Peer Y, Moreau H. (2006) Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. *Proc Natl Acad Sci USA.* 103(31):11647-52.

Devanapally S, Ravikumar S, Jose AM. (2015) Double-stranded RNA made in *C. elegans* neurons can enter the germline and cause transgenerational gene silencing. *Proc Natl Acad Sci USA.* 112(7):2133-8.

Devanapally S, Allgood S, Jose AM. (2017) Mating can cause transgenerational gene silencing. *BioRxiv*. Online on Jun 11.

Dawkins R. (1989) The Selfish Gene. Oxford University Press. p. 192.

DeVries, H. (1906) Species and Varieties: Their origin and mutation. 2nd edition. London: The Open Court Publishing Company. p. 475.

Eiraku M, Takata N, Ishibashi H, Kawada M, Sakakura E, Okuda S, Sekiguchi K, Adachi T, Sasai Y. (2011) Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature.* 472:51–56.

Englander SW, Mayne L. (2014) The nature of protein folding pathways. *Proc Natl Acad Sci USA.* 111(45):15873-80.

Feng S, Jacobsen SE, Reik W. (2010) Epigenetic reprogramming in plant and animal development. *Science.* 330(6004):622-7.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature.* 391(6669):806-11.

Gantz VM, Bier E. (2015) The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science.* 348(6233):442-4.

Gerson-Gurwitz A, Wang S, Sathe S, Green R, Yeo GW, Oegema K, Desai A. (2016) A small RNA-catalytic Argonaute pathway tunes germline transcript levels to ensure embryonic divisions. *Cell.* 165:396–409.

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science.* 329(5987):52-6.

Glass JI, Merryman C, Wise KS, Hutchison CA 3rd, Smith HO. (2017) Minimal Cells-Real and Imagined. *Cold Spring Harb Perspect Biol.* E-pub Mar 27.

Gould SJ, Lewontin RC. (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci* 205:581–598.

Gould SJ, Eldredge N. (1993) Punctuated equilibrium comes of age. *Nature.* 366(6452):223-7.

Grant B, Hirsh D. (1999) Receptor-mediated endocytosis in the Caenorhabditis elegans oocyte. *Mol Biol Cell.* 10(12):4311-26.

Greer EL, Maures TJ, Ucar D, Hauswirth AG, Mancini E, Lim JP, Benayoun BA, Shi Y, Brunet A. (2011) Transgenerational epigenetic inheritance of longevity in Caenorhabditis elegans. *Nature.* 479(7373):365-71.

Gurdon JB, Elsdale TR, Fischberg M. (1958) Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. *Nature.* 182(4627):64-5.

Gustafsson A. (1979) Linnaeus' peloria: the history of a monster. *Theor. Appl. Genet.* 54:241-8.

Halfmann R, Lindquist S. (2010) Epigenetics in the extreme: prions and the inheritance of environmentally acquired traits. *Science.* 330(6004):629-32.

Heinze J, Schrempf A. (2008) Aging and reproduction in social insects--a mini-review. *Gerontology.* 54(3):160-7.

Helm M, Motorin Y. (2017) Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat Rev Genet.* 18(5):275-291.

Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA. (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol.* 33(5):510-7.

Hollick JB. (2017) Paramutation and related phenomena in diverse species. *Nat Rev Genet.* 18(1):5-23.

Hubbard EJA, Greenstein D. (2015) Introduction to the germ line, *WormBook*, ed. The *C. elegans* Research Community., WormBook, doi/10.1895/wormbook.1.18.1, http://www.wormbook.org.

Hutchison CA 3rd, Chuang RY, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi ZQ, Richter RA, Strychalski EA, Sun L, Suzuki Y, Tsvetanova B, Wise KS, Smith HO, Glass JI, Merryman C, Gibson DG, Venter JC. (2016) Design and synthesis of a minimal bacterial genome. *Science.* 351(6280):aad6253.

Huxley J. (2010) Evolution: The Modern Synthesis – the definitive edition. Cambridge: MIT Press (originally published in 1942 London: Allen and Unwin). p. 555.

Jablonka E, Lamb ML. (2005) Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life. Cambridge: The MIT Press.

Jablonka E, Raz G. (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Quant Rev Biol.* 84(2):131-76.

Jacob F. (1977) Evolution and tinkering. *Science.* 196:1161-6.

Jacob F, Monod J. (1961) Gene regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3: 318-56.

Kauffman S. (1969) Homeostasis and differentiation in random genetic control networks. *Nature.* 224(5215):177-8.

Klosin A, Casas E, Hidalgo-Carcedo C, Vavouri T, Lehner B. (2017) Transgenerational transmission of environmental information in *C. elegans*. *Science.* 356(6335):320-323.

Komor AC, Badran AH, Liu DR. (2017) CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell.* 168(1-2):20-36.

Lancaster MA, Renner M, Martin CA, Wenzel D, Bicknell LS, Hurles ME, Homfray T, Penninger JM, Jackson AP, Knoblich JA. (2013) Cerebral organoids model human brain development and microcephaly. *Nature.* 501(7467):373-9.

Lassar AB, Paterson BM, Weintraub H. (1986) Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. *Cell.* 47(5):649-56.

Leopold LE, Heestand BN, Seong S, Shtessel L, Ahmed S. (2015) Lack of pairing during meiosis triggers multigenerational transgene silencing in Caenorhabditis elegans. *Proc Natl Acad Sci USA.* 112(20):E2667-76.

Lev I, Seroussi U, Gingold H, Bril R, Anava S, Rechavi O. (2017) MET-2-Dependent H3K9 Methylation Suppresses Transgenerational Small RNA Inheritance. *Curr Biol.* 27(8):1138-1147.

Levine M, Davidson EH. (2005) Gene regulatory networks for development. *Proc Natl Acad Sci USA.* 102(14):4936-42.

Lillie FR. (1896) On the smallest parts of Stentor capable of regeneration. *J. Morphol.* 12:239-249.

Linnaeus C, Rudberg D. (1744) De Peloria. Diss. Ac. Amoenitates Academicae III, Uppsala.

Luteijn MJ, van Bergeijk P, Kaaij LJ, Almeida MV, Roovers EF, Berezikov E, Ketting RF. (2012) Extremely stable Piwi-induced gene silencing in Caenorhabditis elegans. *EMBO J.* 31(16):3422-30.

Lynch M. (2007) The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci USA.* 104 Suppl 1:8597-604.

Madhani HD. (2013) The frustrated gene: origins of eukaryotic gene expression. *Cell.* 155(4):744-9.

Marshall W. (2015) Ten craziest things cells do. https://www.ibiology.org/ibioeducation/exploring-biology/cell-bio/signaling/ten-craziest-things-cells-do.html.

Marré J, Traver EC, Jose AM. (2016) Extracellular RNA is transported from one generation to the next in *C. elegans*. *Proc. Natl. Acad. Sci. USA.* 113(14):12496-501.

Mayr E. (1961) Cause and effect in biology. Science. 134:1501-6.

McLaren A. (1980) Germ cells and soma: a new look at an old problem. New Haven and London: Yale University Press. p. 3.

Mereschkowsky C. (1905). Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol. Centralbl.,* 25: 593–604. English translation in Martin, W., Kowallik, K. V. (1999). Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren im Pflanzenreiche'. *Eur. J. Phycol.,* 34: 287–295.

Monod J, Jacob F. (1961) Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harbor Symp. Quant. Biol.* 26: 389–401.

Murray A, Hunt T. (1993) The Cell Cycle: An Introduction. 1st edition. New York: Oxford University Press.

Nanney DL. (1958) Epigenetic control systems. *Proc Natl Acad Sci USA.* 44(7):712-7.

Nirenberg NW, Matthaei JH. (1961) The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci USA.* 47:1588-602.

Olsen OW. (1986) Animal Parasites: Their Life Cycles and Ecology. Dover Publications.

Ostrov N, Landon M, Guell M, Kuznetsov G, Teramoto J, Cervantes N, Zhou M, Singh K, Napolitano MG, Moosburner M, Shrock E, Pruitt BW, Conway N, Goodman DB, Gardner CL, Tyree G, Gonzales A, Wanner BL, Norville JE, Lajoie MJ, Church GM. (2016) Design, synthesis, and testing toward a 57-codon genome. *Science.* 353(6301):819-22.

Park M, Keung AJ, Khalil AS. (2016) The epigenome: the next substrate for engineering. *Genome Biol.* 17(1):183.

Phillips EF. (1903) A Review of Parthenogenesis. *Proc. Amer. Phil. Soc.* 42(174):275-345.

Pisco AO, d'Herouel AF, Huang S. (2016) Conceptual confusion: the case of epigenetics. *BioRxiv.* Online on May 12.

Rafelski SM, Marshall WF. (2008) Building the cell: design principles of cellular architecture. *Nat Rev Mol Cell Biol.* 9(8):593-602.

Rankin CH. (2015) A review of transgenerational epigenetics for RNAi, longevity, germline maintenance and olfactory imprinting in Caenorhabditis elegans. *J Exp Biol.* 218(Pt 1):41-9.

Rechavi O, Houri-Ze'evi L, Anava S, Goh WS, Kerk SY, Hannon GJ, Hobert O. (2014) Starvation-induced transgenerational inheritance of small RNAs in C. elegans. *Cell.* 158(2):277-87.

Rivoire O, Leibler S. (2014) A model for the generation and transmission of variations in evolution. *Proc Natl Acad Sci USA.* 111(19):E1940-9.

Rodgers AB, Morgan CP, Bronson SL, Revello S, Bale TL (2013) Paternal stress exposure alters sperm microRNA content and reprograms offspring HPA stress axis regulation. *J Neurosci.* 33(21):9003–9012.

Sagan L. (1967) On the origin of mitosing cells. *J Theoret Biol.* 14:255–274.

Salmon ED. (1975) Pressure-induced depolymerization of spindle microtubules. I. Changes in birefringence and spindle length. *J Cell Biol.* 65: 603-614.

Sander K. (1991) Wilhelm Roux and his programme for developmental biology. *Roux's Arch Dev Biol.* 200:1-3.

Schwille P. (2015) Jump-starting life? Fundamental aspects of synthetic biology. *J Cell Biol.* 210(5):687-90.

Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, Belleannee C, Kucukural A, Serra RW, Sun F, Song L, Carone BR, Ricci EP, Li XZ, Fauquier L, Moore MJ, Sullivan R, Mello CC, Garber M, Rando OJ. (2016) Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science.* 2016 351(6271):391-6.

Shirayama M, Seth M, Lee HC, Gu W, Ishidate T, Conte D Jr, Mello CC. (2012) piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. *Cell.* 150(1):65-77.

Siklenka K, Erkek S, Godmann M, Lambrot R, McGraw S, Lafleur C, Cohen T, Xia J, Suderman M, Hallett M, Trasler J, Peters AH, Kimmins S. (2015) Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science.* 350(6261):aab2006.

Sood AJ, Viner C, Hoffman MM. (2016) DNAmod: the DNA modification database. *BioRxiv*. Online on Aug 26.

Tachibana M, Sparman M, Sritanaudomchai H, Ma H, Clepper L, Woodward J, Li Y, Ramsey C, Kolotushkina O, Mitalipov S. (2009) Mitochondrial gene replacement in primate offspring and embryonic stem cells. *Nature.* 461(7262):367-72. Erratum in: *Nature.* 516(7530):276.

Takahashi K, Yamanaka S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* 126(4):663-76.

Thanbichler M, Shapiro L. (2008) Getting organized--how bacterial cells move proteins and DNA. *Nat Rev Microbiol.* 6(1):28-40.

Thompson EL. (1951) Time and rate of loss of nuclei by the red blood cells of human embryos. *Anat Rec.* 111(3):317-25.

Thul PJ et al. (2017) A subcellular map of the human proteome. *Science.* 356(6340). 10.1126/science.aal3321.

Tsioris K, Torres AJ, Douce TB, Love JC. (2014) A new toolbox for assessing single cells. *Annu Rev Chem Biomol Eng.* 5:455-77.

de Vanssay A, Bougé AL, Boivin A, Hermant C, Teysset L, Delmarre V, Antoniewski C, Ronsseray S. (2012) Paramutation in Drosophila linked to emergence of a piRNA-producing locus. *Nature.* 490(7418):112-5.

Vastenhouw NL, Brunschwig K, Okihara KL, Müller F, Tijsterman M, Plasterk RH. (2006) Gene expression: long-term gene silencing by RNAi. *Nature.* 442(7105):882.

Waddington CH. (1942) The Epigenotype. *Endeavour* 1:18-20.

Waddington CH. (1957) The strategy of the genes. New York: George Allen & Unwin Ltd. p. 32.

Waddington CH. (1962) New Patterns in Genetics and Development. Columbia University Press.

Waddington CH. (1968) Towards a Theoretical Biology, Vol. 1, Prolegomena. Edinburgh: International Union of Biological Sciences & Edinburgh University Press. pp. 1–32.

Weiss SB, Gladstone L. (1959) A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. *J. Am. Chem. Soc.* 81: 4118-4119.

Wilson EB. (1896) The Cell in Development and Inheritance. New York: The Macmillian Company.

Wolke U, Jezuit EA, Priess JR. (2007) Actin-dependent cytoplasmic streaming in *C. elegans* oogenesis. *Development.* 134(12):2227-36.