Stochasticity or the fatal 'imperfection' of cloning

REINER A VEITIA

Université Denis Diderot/Paris VII
INSERM U361 Reproduction et Physiopathologie Obstetricale,
Hospital Cochin, Pavillon Baudelocque, 123 Bd de Port Royal, 75014 Paris, France

(Fax, 33-1-43264408; Email, veitia@cochin.inserm.fr)

The concept of clone is analysed with the aim of exploring the limits to which a phenotype can be said to be determined geneticaly. First of all, mutations that result from the replication, topological manipulation or lesion of DNA introduce a source of heritable variation in an otherwise identical genetic background. But more important, stochastic effects in many biological processes may superimpose a phenotypic variation which is not encoded in the genome. The source of stochasticity ranges from the random selection of alleles or whole chromosomes to be expressed in small cell populations, to fluctuations in processes such as gene expression, due to limiting amounts of the players involved. The picture emerging is that the term clone is a statistical over-simplification representing a series of individuals having essentially the same genome but capable of exhibiting wide phenotypic variation. Finally, to what extent fluctuations in biological processes, usually thought of as noise, are in fact signal is also discussed.

1. Introduction

What is a clone? This seemingly simple question entails more complexity that one might suspect at first sight. Even a simplistic answer may be produced at different levels. At the level of DNA sequence, a clone can be defined as a group of individuals sharing the same genomic sequence. But is that possible? The answer is not straightforward, as manipulation of the genetic information by the cellular machinery or environmental DNA-damaging factors leads to the introduction of mutations. To see this, let us start with a single cell of the bacterium *Escherichia coli* growing to form a colony. We will consider that the mutation rate is of the order of 10^{-10} per base pair (pb), in the presence of postreplicative repair systems, and that *E. coli* contains some 4,000 genes of 1,000 bp each (Blattner *et al* 1997). Under these assumptions, mutations will arise on average in a fraction of 4×10^{-4} per generation.

That is, after 1 generation, 0·9996 of the population remains 'wild-type' (ignoring lethal mutations). Thus, after n generations, the proportion of wild-type bacteria left within the 'clone' will be $0·9996^{n}$. After 260 generations about 10% of the individuals will be mutants (re-estimated from Davey and Kell 1996). This takes about 4–5 days. If we are too strict, in less than a week, the *E. coli* mutants will have shattered the concept of clone at the genetic level. If we are permissive, and we want the concept of clone to survive, we should accept that a true perfect clone from a genomic point of view is an over-simplification. If we isolate the 10% of mutants we just talked about, we will probably detect large variations in their transcriptomes. Yet, these mutants differ from each other in genetic terms by only 0·2 ppm (parts per million)! Next, we try to conduct this exploration beyond DNA. We will discuss how the same genotype may lead to 'non-coded' phenotypic variation.

Keywords. Clone; mutation; monoallelic expression; stochasticity.

2.   Monoallelic expression in small cell populations

As far as we know, most genes in diploids are expressed from both alleles. However, there is a growing class of genes transcribed preferentially from a single allele in each cell. Three main mechanisms can explain monoallelic expression. (i) In humans and other mammals, males have one copy of the X chromosome, whereas normal females have two copies. This potential imbalance of X-linked gene dosage is circumvented by inactivating one X- chromosome in the female (Willard 1996). (ii) In a more subtle way, genomic imprinting is a process leading to expression of either the maternal or paternal allele in a particular locus. This previous definition concerns strict imprinting, which implies expression of the same allele in all the relevant cells. However, there are more relaxed cases where imprinting is limited to certain tissues or developmental stages or even to certain individuals in the population (polymorphic imprinting) (Weinstein 2001). (iii) Finally, allelic exclusion is another category of monoallelic expression where the choice of the allele to be expressed (maternal or paternal) is random from one cell to another (Nutt and Busslinger 1999). There are several ways to explain the stochastic choice of one chromosome for X-inactivation or one allele or the other in the case of allelic exclusion. Possibly, *trans*-acting factors, which bind the *cis*-regulatory elements that control expression, are present in limiting amounts only sufficient to bind one allele at a time (Brown and Chandra 1973). If this is so, the initial binding event must leave a permanent mark turning on a single *cis*-controlling element per cell. There may also be allele-to-allele cross talk, whose exact nature is still unclear. Thus, in the nucleus, the two alleles may co-localize, allowing for *trans*-allelic interactions as proposed for the X-inactivation centre. Whatever the mechanism, the result is differential DNA methylation, asynchronous DNA replication, differential chromatin modifications and unequal nuclear localization (Ogawa *et al* 2003).

Imagine now that we study a very large cell population in which the random decision to express one allele or the other from a locus is to be made. Clearly this is an old laplacian probability problem, much like tossing a coin. Fifty percent of the cells will express one allele and 50% will express the other. The outcome of this process is deterministic. However, things can be very differently if the starting population is small. In such a small cell population, the outcome critically depends on the initial number of cells in which the choice is to be made. The smaller it is the larger the departures from 50 : 50 will be. For instance, assume that the initial cell population, that will form a tissue after differentiation, contains 20 cells and that the probability of inactivating either allele is 0·5.

That is, allele inactivation is really random. After exploring many cell populations, it is clear that the mean per centage of inactivation is 50 : 50 but large excursions in the proportions are noticed from one cell population (i.e. in one individual) to another (figure 1). A similar phenomenon takes place during the process of X-inactivation. One can explore allele or X-inactivation issues using the binomial probability law. This will allow us to calculate the proportion of cells expressing each type of allele or chromosome, starting from a population of any size, assuming absence of selection advantages linked to any of the relevant alleles or chromosomes. Thus, using the binomial law it is possible to estimate the number of cells where X-inactivation occurred, for instance, in the lymphocyte precursors. This kind of analysis has been performed with the observed data from young Safari cats. Estimates suggest that approximately 16 cells were present at the time of X-chromosome inactivation, or alternatively that 16 cells gave rise to the hematopoietic system during development (Abkowitz *et al* 1998). In the human female, X-chromosome inactivation is a stochastic event that occurs early in embryonic development, around the 100-cell stage (Lyon 1962). However, according to our previous arguments, it is very likely that in tissues arising from cell subpopulations, much smaller than 100, there will be normal skews in X-inactivation. This may explain the widespread phenotypic discordance between monozygotic (genetically identical) twins as discussed below.

For the sake of argument, imagine a species where XX females can reproduce by parthenogenesis under certain conditions. This will lead to a clonal population. We think automatically that the mother and the daughters will be identical. This is almost certainly so at the genotypic level, if we disregard the rare mutational events we just talked about. But what happens at the phenotypic level? A hint of the answer comes from the example outlined above. In fact, it all depends on the number of cells in which the choice of various monoallelically expressed genes or even X-inactivation took place during ontogenesis. Interestingly, in bilaterally symmetric organisms, there is often a fluctuating asymmetry; that is, random differences exist between left and right sides. However, both sides share the same genome and develop under similar environmental conditions. Therefore, the variation of symmetry around an average value is suspected to be due to fluctuations of developmental processes (Klingenberg 2003). A straightforward source of asymmetry comes from the stochastic choice of monoallelically expressed developmental genes from limited primordial cell populations. The potential relevance of this seemingly biologically hazardous strategy (i.e. random choice in small populations) will be discussed later.

3.   Stochasticity in cellular processes

Much of the variation of gene expression in time, space or both has been attributed to stochastic events. When the factors that connect a gene with its product in the information transfer chain are present at very low concentrations, gene expression becomes a probabilistic process. Under such conditions, the amount of gene product produced by a cell is the result of the integration of small quantal events (McAdams and Arkin 1997). They involve a series of discontinuous processes such as promoter recognition, effective transcription of the gene, RNA splicing if required and finally, translation, in the case of proteins. The rates at which each of these steps take place depends to a great extent on the concentration of the relevant actors and their probability of forming productive complexes. Specifically, during transcription, the probability of promoter recognition is proportional to the concentration of transcription factor (TF) (Boulanger *et al* 1987; Kato *et al* 1986). At very low concentration of a TF, fluctuations in reaction rates cause large random variation leading to a dynamic equilibrium between switching *on* and *off* expression. At high TF concentrations, the promoter is virtually saturated and the discontinuous process of transcription may take the appearance of a fairly continuous process. This is analogous to the functioning of a com-bustion motor in a car. A motor with only one cylinder will work jerkily while another with six cylinders will give the appearance of working rather continuously even if the underlying mechanism can be decomposed into discrete phases.

Examples of stochastic gene expression in clonal prokaryotes have been known since a long time. For instance, over 40 years ago Novick and Weiner (1957) reported that enzyme induction from the lactose (lac) operon in *E. coli* was an all-or-none phenomenon. At sub-saturating levels of the inducer substrate, they observed co-existence of lac expressing as well as silent cells. A plausible explanation runs as follows. Repression of the lac promoter is tight but from time to time, in the absence of inducer, there are random bursts of transcription (leakiness of the promoter) leading to the synthesis of the membrane transporter for lactose. At low concentrations of inducer, only the fraction of cells containing enough pre-synthesized transporter (lac permease) will de-repress the operon. With permease accumulation, these individual cells and their offspring become stably de-repressed as a result of this positive feedback loop. Concomitantly, there is an unresponsive sub-population where the threshold concentration of permease is not attained. A similar phenomenon may explain the behaviour of the regulon (i.e. set of co-regulated operons) controlling arabinose
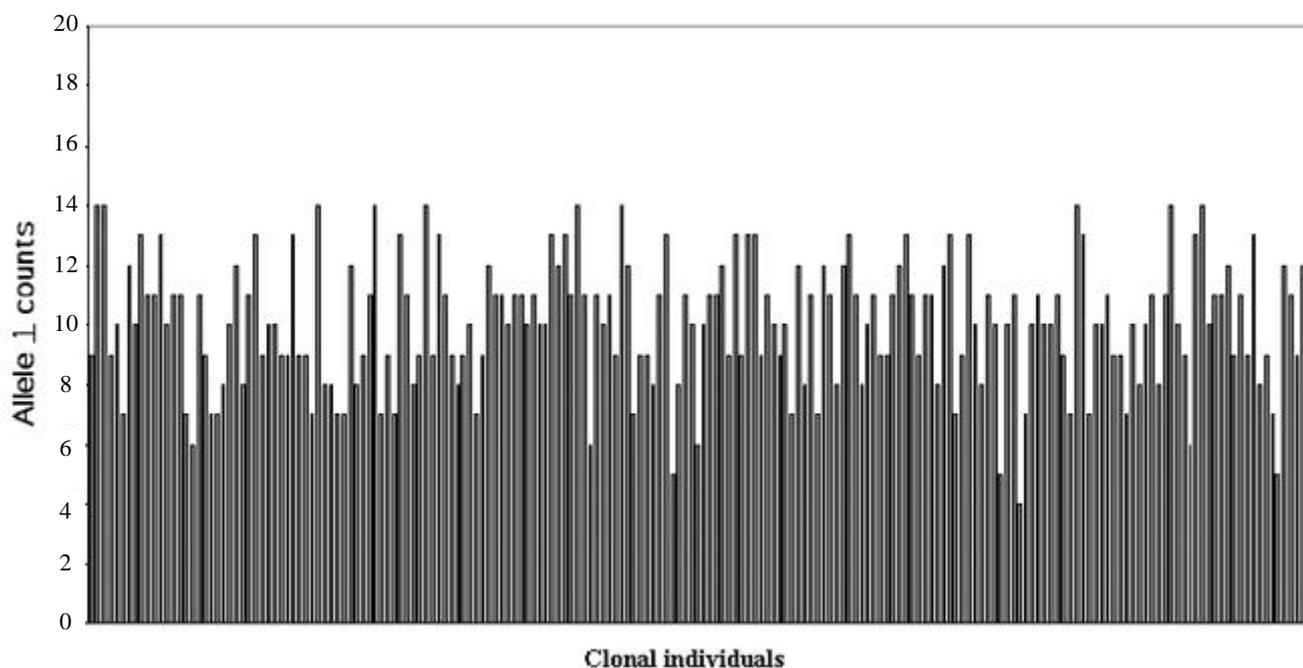

Figure 1.   Representation of a binomial process in small cell populations in clonal individuals. Two alleles (0 and 1) are chosen at random to be expressed in populations of 20 cells. The graph represents the number of cells expressing the allele of type 1, in 100 different populations. The populations in the graph have not been ranked in any way (i.e. in the form of a histogram), to show how large the fluctuations around 10 (the mean) can be.

utilization. Induction of the *ara* regulon relies on the interaction between arabinose and the transactivator AraC, which controls the synthesis of two arabinose uptake systems. As above, following stochastic bursts of expression of the *ara* regulon, there may be a fraction of the cells expressing the *ara* uptake proteins (among other genes). As in the case of the lac permease, the concentration of the transporters will decrease in the expressing cells as the proteins are diluted by cell growth, and the probability that individual cells will accumulate inducing levels of arabinose before losing the transporters will depend on the concentration of arabinose (Siegele and Hu 1997). This leads to a sigmoidal dose-response curve. Stochasticity in the case of de-repression of the lac operon has recently been re-demonstrated using bacterial clones able to express two distinguishable fluorescent proteins whose transcription is driven by the lac promoter. At intermediate levels of the inducer, relatively random and intermittent expression of each fluorescent protein is detected (Elowitz *et al* 2002). Gardner *et al* (2002) have created an engineered genetic toggle with two mutually repressible promoters, one of which also drives the expression of the green fluorescent protein (GFP). This system has two states (low GFP or high GFP). Gardner *et al* (2002) studied the behaviour of the system with deterministic equations. This analysis predicts a discontinuous jump from one state to the other. However, the stochastic nature of gene expression causes variability in the location of the switching threshold and thus blurs the point of the jump, where GFP-expressing and silent cells co-exist (figure 2). These few examples coming from bacteria, a simple form of life, illustrate how members of a clonal population that can be identical at the genomic level can display an all-or-none response with regard to a certain phenotype. It is interesting to speculate what the picture would look like if we were able to see several phenotypes at the same time. Probably all the cells within the population would behave differently from each other.

Many variables beyond transcription initiation affect gene expression. They include the efficiency of transcriptional elongation, the rate of splicing, the half-life of mRNA, the efficiency of translation and the half-life of the protein. In a bacterial model using a GFP reporter, translation efficiency seems to be the major factor explaining variation in GFP levels (Ozbudak *et al* 2002). However, recent work in yeast (Blake *et al* 2003) has shown that the major determinant of GFP reporter fluctuations were stochastic variations in mRNA levels. In both systems, higher translational efficiencies were associated with higher levels of noise, suggesting that high efficiency of translation can amplify fluctuations experienced at the transcriptional level. In fact, efficient translation would produce a more direct correlation between fluctuations in mRNA levels and protein levels. In con-

trast, a large number of RNA molecules would provide a buffer against noise associated with inefficient translation. The latter case has a mechanical analogy. The axis of the motor of a car is coupled to a very heavy metal disk, which plays the role of an accumulator of kinetic energy. The high inertia of the disks buffers the underlying jerky functioning of the motor.

Stochastic gene expression is not limited to prokaryotes. In eukaryotic transcription also, a dynamic equilibrium between the *on* and *off* states of gene activity is expected. A few direct measurements of the kinetics of gene expression have been performed. For instance, the analysis of several globin genes indicates that expression of each gene is intermittent (Wijgerde *et al* 1995) and some tran-
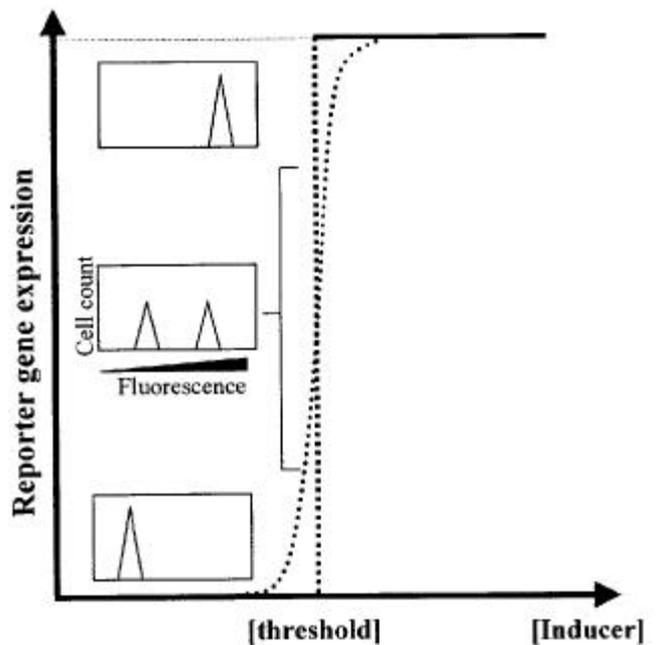

Figure 2. Properties of an engineered genetic toggle with two mutually repressible promoters, one of which drives also the expression of the GFP (Gardner *et al* 2002). This system has two states (low GFP or high GFP) elicited according to particular inductive signals. The behaviour of the system was studied with deterministic equations. This analysis predicts a discontinuous jump from one state to the other, as shown in the graph (inducer concentration vs reporter GFP expression). This behaviour is of course difficult to conceive in biology. Accordingly, the stochastic nature of gene expression causes variability in the location of the switching threshold and thus blurs the point of jump. The small rectangular inserts, represent schematically the output of the fluorescent activated cell sorter. The peaks represent cell populations expressing (right) or not (left) the GFP. Near the bifurcation, there is a bimodal distribution of cells. One outstanding property of this system is that it provides the basis a conditional epigenetic memory. That is, after removing the inductive signal the system remains in the state where it was and this character is transmitted to the next generations.

scription complexes seem to alternate between bound and unbound states. The binding kinetics of the estrogen receptor to a regulatory region containing several estrogen responsive elements depends on ligand concentration (Shang *et al* 2000). Increasing the latter increases the probability of turning the gene on and the average amount of time the gene is transcribed. Even more striking, it has recently been shown that most active genes undergo *on-off* transcription cycles and can co-localize with nuclear subcompartments that concentrate RNApol II (i.e. transcriptional factories). *In situ* hybridization experiments on nascent DNA clearly show that even for highly expressed genes the two alleles of each loci are not expressed in all cells analysed. The *on* state seems to be linked to transcription factory occupancy and the *off* state with repositioning away from the factories. Gene induction is expected to occur through a rapid and transient increase in the number of alleles localized in transcription foci within the cell population. These findings substantiate the notion that genes dynamically move in and out of transcription factories, resulting in activation and abatement of their transcription (Osborne *et al* 2004). Even in the context of multinucleate cells, as in skeletal muscle fibers, the same gene is not transcriptionally active at a given moment in all the nuclei and the activity of each locus is regulated independently of the others. Transcription occurs in pulses and is defined by a stochastic mechanism. In fact, the total number of active loci for a particular gene is dynamic, changing during foetal development and in the adult (Newlands *et al* 1998).

It is widely accepted that certain types of transcriptional enhancers work stochastically by increasing the probability of a promoter of becoming active (Fiering *et al* 2000). A distinct characteristic of this binary mode of enhancer action is that in a cellular population not all cells will express the gene stimulated by the enhancer. This is very likely due to differences in the enhancer occupancies by the relevant transcription factors, among different cells, that lead to stable changes. A functional enhancer may also antagonize gene silencing by preventing localization of a gene near heterochromatin. To explore this mechanism Francastel *et al* (1999) have assessed the ability of an enhancer and its mutants to influence silencing and nuclear location of a transgene. Disruption of core enhancer motifs impairs the enhancer's anti-silencing capacity. Accordingly, transgenes linked to the functional enhancer localize away from centromeric heterochromatin in the interphase nucleus while mutations in the enhancer result in increased rates of transgene silencing as it localizes close to the centromeric heterochromatin. It is known that when an euchromatic gene is juxtaposed with a heterochromatic domain, it may be subjected to variable, but clonally heritable, repression. This binary/stochastic phenomenon is called position effect variegation

(PEV) and is related to other phenomena of transcriptional silencing that are inherited by epigenetic mechanisms. PEV was discovered by Muller (1930) who described mutations of the *white +* (*w +*) eye colour gene of *Drosophila* resulting in huge cell-to-cell variations in gene expression. The phenotype is due to a position effect in which a chromosomal rearrangement breakpoint put the *w +* gene in the vicinity of heterochromatin. This led to big patches of red facets adjacent to big white patches in the adult eye, suggesting that a decision to express or repress the *w +* gene was made early in development and was maintained through multiple cell divisions (for review, see Wakimoto 1998). Budding yeast heterochromatin can also repress adjacent genes. When a marker gene is integrated near the telomere it can be repressed epigenetically. This telomere position effect is similar to PEV (see Grunstein 1998). Indeed, PEV itself has also been detected on genes located near the mating-type locus of fission yeast (Ayoub *et al* 1999). PEV shows that the boundary between active euchromatin and heterochromatin is labile at least at certain moments. Some enhancers might help to better define this rather dynamic boundary. For instance, Sutherland *et al* (1997) have examined the effect of a globin enhancer on the variegation of *b*-galactosidase gene expression in erythrocytes of transgenic mice. Mice carrying the *b*-galactosidase gene driven by the *a*-globin promoter, without the enhancer, exhibited enzyme expression in a very small proportion of embryonic erythrocytes. When the construct also contained a specific enhancer, expression was detected in a much higher proportion of embryonic erythrocytes. A decline in transgene expression occurred as mice aged, mainly due to a decrease in the proportion of cells expressing the transgene. Interestingly, in the post-natal life there was a fairly exponential decay of the proportion of expressing cells (figure 3). This is reminiscent of a random process characterized by a constant extinction rate, as in radioactive decay. Accordingly, one cannot predict when a particular precursor cell will stop producing the reporter *b*-galactosidase. Predictions are only valid for the ensemble, that will behave deterministically (i.e. exponential extinction) if the number of cells analysed is high.

Binary behaviour of eukaryotic gene expression can also result from de-repression, as it has been shown for the yeast Gal1 promoter (Biggar and Crabtree 2001). This phenomenon is observed when haploid cells are grown in glucose (which provokes repression) and then transferred to an inducing medium (galactose) with varying concentrations of glucose. Gal80, a repressor of the transcription factor Gal4, mediates this response. Changes in the medium induce modifications of Gal80 that make it lose its capacity to repress Gal4, which, in turn, activates the promoter. A clear sigmoidal relationship between glucose concen-

tration (responsible for repression) and the proportion of cells where the promoter is active is observed (figure 4; see also Veitia 2003a).

To gain insights into binary/stochastic behaviour in eukaryotes, Becskei *et al* (2001) created a synthetic gene switch in yeast involving a positive feedback loop. This simple system was able to transform a constitutively expressed transcriptional activator, sensitive to an inducer (doxycycline), into a binary switch when the transactivator was expressed autocatalytically. As in the case of the Gal1 promoter responding to glucose + galactose, a continuous input gave rise to a bimodal distribution of cells

Figure 3. A globin enhacer can affect variegation of expression of the reporter **b**-galactosidase gene in fetal erythrocytes of transgenic mice. With age, there is a decline in transgene expression. The graphic represents the percentage of **b**-galactosidase expressing cells versus the age of mice. In the post-natal life there is a fairly exponential decay of the proportion of expressing cells (notice the semi-log character of the graph). This is reminiscent of a random process characterized by a constant extinction rate, as that of radioactive decay. Under this assumption an exponential decay law is obtained (mathematical derivation in Schweidler 1906). According to this model, the probability of extinction of expression in the erythroid precursors would be independent of time and of the age of a particular cell. Thus, one cannot predict when a particular precursor cell will stop expressing the reporter gene. Predictions are only valid for the ensemble, that will behave deterministically (i.e. exponential extinction of expression) if the number of cells studied is high. Reproduced and modified with permission from the Journals Department, American Society for Microbiology and from the author (E Whithelaw).

(responding and silent). Moreover, the proportion of cells displaying either low or high expression states varied with changes in the inducer concentration.

4. Stochasticity in development

During pattern formation in development, medium- and long-distance regulatory signals are often distributed as a gradient that is translated into more or less sharp boundaries (Lewis *et al* 1977; Murray 1989). However, even the sharpest boundaries involve cells with different phenotypes. For instance, Nijhout (2005) has proposed that stochastic gene expression underlies development of the vivid colour pattern of butterflies made up of pigmented scales. Each scale is the outgrowth of a single epidermal cell and typically contains a single pigment due to an all-or-none decision to synthesize alternative pigments in response to a signal. However, at the boundary between regions with well-defined colours, there is a salt-pepper pixellated mosaic of monochrome scales. The existence of stochastic events in development implies that the values of a developmental parameter and the associated phenotype are not fully determined by the genotype and the environment. In fact, there is variation around the expected parameter value for a given genotype in a given environment. The phenotypic outcome of a small fluctuation in the context of a developmental process depends on the shape of the function that relates the genotype to the phenotype. The latter is called the 'mapping function' (see Klingenberg 2003). If
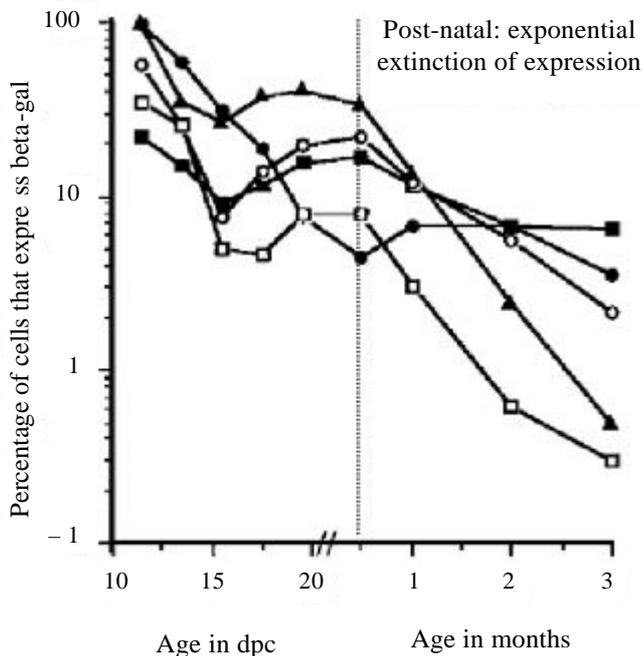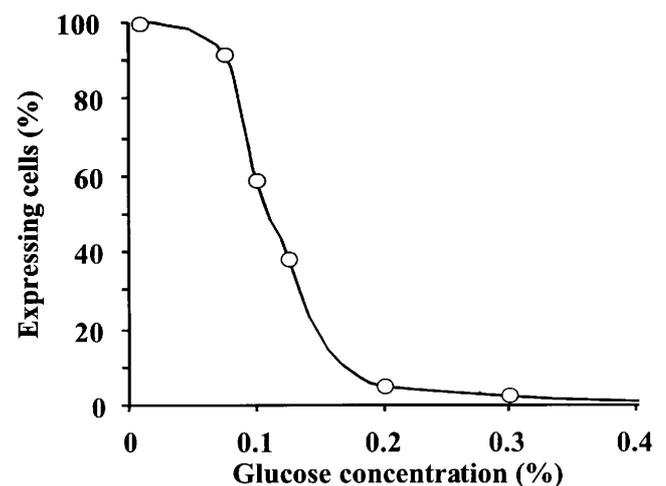

Figure 4. Sigmoidal response to de-repression. The yeast Gal1 promoter can display a binary response when haploid cells are grown in 2% glucose (repressor) and then transferred to an inducing medium (2% galactose) with varying concentrations of glucose (as indicated in the graph) for 14 h. Gal-responding cells display similar levels of transcription.

the mapping function is steep, a small developmental perturbation will have a large effect. On the other hand, a flatter mapping function will ensure a higher developmental stability (figure 5).

Developmental fluctuations may arise from the stochasticity inherent in the biochemical process of gene expression or in the choice of the genes to be expressed, as well as from fluctuations in cellular components or other parameters such as concentration of hormones, growth factors and temperature. These stochastic processes may produce observable morphological variation (Klingenberg 2003). As mentioned above, fluctuations may lead to a developmental instability that can translate into asymmetry in bilaterally symmetric organisms. This asymmetry is a function of opposing forces: the fluctuations themselves and the buffer capacity of the system to counter them. In a recent work, Kellner and Alford (2003) detected substantial asymmetry for four bilateral characters during the development of domestic fowl. Asymmetry was random in degree and direction and decreased rapidly with developmental time, probably due to some kind of feedback. Indeed, feedback mechanisms are important in consolidating developmental decisions. For
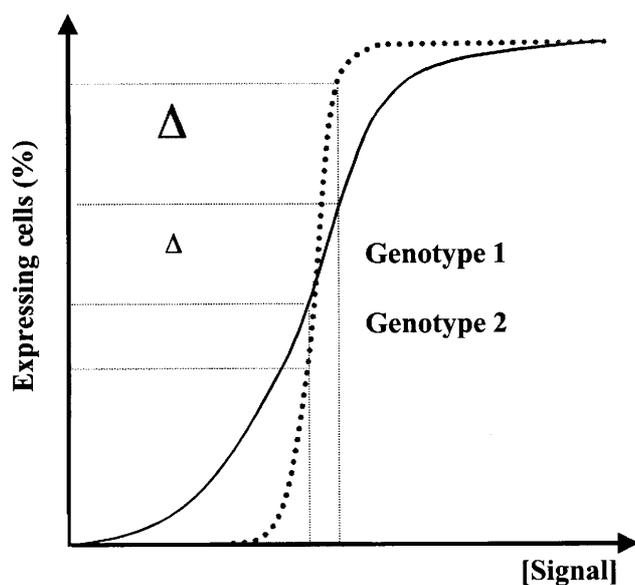
instance, the community effect is a mechanism to establish boundaries among different domains during development. The activation of a gene in a particular cell leads to the production of a signalling factor (diffusible or propagated by cell to cell contacts) that stimulates the expression of the same gene in the original cell (positive feedback) and its responsive neighbours. Activation of this gene in one of the cells of the community, even to a low extent, will trigger the gene in all the cells, after a delay determined by the time required for the transmission of the signal from cell to cell (Gurdon *et al* 1993). Such a developmental process may rely on a stochastic event (expression in the triggering cell) that subsequently leads to a deterministic outcome (successful tissue determination).

A heterozygous loss-of-function mutation can lead to a disease or very abnormal phenotypes (haploinsufficiency). It has been proposed that stochastic developmental effects may explain the dramatic results of haploinsufficiency (Cook *et al* 1998) that lead, in general, to a striking variability from one individual to another bearing the same mutations or even within the same individual (bilateral asymmetry). For example, in a hermaphrodite, there may be a testis on one side and an ovary at the other side of the body. The existence of a developmental step operating extremely close to a threshold in pathological conditions may account for this variability. It is more and more apparent that the fate of a tissue is not always determined by the expression of a set of genes above a certain threshold in all cells but rather by the proportion or density, of expressing cells. In a haploinsufficient condition in a multicellular organism, it is possible that when only one allele of the locus is working, the concentration of the particular factor may be insufficient to exert its role in all cells. However, at the level of the cell population, there may be some cells where this has been possible. Crossing a critical threshold for normal development to occur is required; otherwise, the developmental pathway will be aborted. Near-threshold levels of essential gene products may lead stochastically to opposite outcomes in the same individual. In this case, the genetic background cannot be used as a possible explanation. Local differences may exert an effect on the outcome (figure 5). Mouse is by far less dosage sensitive than man (Veitia 2003a). That is, the mouse orthologs of many human haploinsufficient genes seem to work perfectly in the heterozygous state. This may be just a reflection of our inability to detect subtle phenotypic changes. However, in the case of genes encoding TFs one is tempted to think of a general mechanism to account for these puzzling differences. What if the interphase chromatin fine structure of human and mouse were different? For instance, it is known that mouse chromosomes are acrocentric and of similar size (Liyanage *et al* 1996). Add to this the extensive shuffling

Figure 5. Graph representing the response of a cell population to a signal S. Two different dose-response curves have been depicted, corresponding to two different genotypes. Note how, for very small changes in S, genotype 1 (inclined sigmoid) responds with small variations while genotype 2 (steep sigmoid) responds with strong changes. For example, in the case of the boundaries of colour domains in the butterfly wings, the sharpness of the boundaries defined by the two curves will be very different. In the case of a haploinsufficient condition where the signal is halved, much larger pehotypic variations are expected for individuals with the genotype 2.

that has taken place in the human and mouse genomes, in some cases keeping but in many others breaking, gene order and location (Ehrlich *et al* 1997). Stemming from these differences, the occupancy of the transcriptional factories in the conditions of haploinsufficiency of a given TF might also be different. Then, uncovered sources of non-linearity would favour occupancy of the transcriptional factories in mouse nuclei. The implication is that with 50% of TF, the transcription of the genes regulated by the TF about 50% of the wild type level.

5. Is stochasticity a dangerous game? Is noise true noise or signal?

There are ways to buffer stochastic fluctuations in gene expression and development. The existence of redundant genes and alternative pathways may help buffer fluctuations, even complete gene deletions. In yeast, the phenotype of about a quarter of full gene deletions is compensated by the corresponding paralogs (Gu *et al* 2003). There are general agents that buffer developmental processes. For instance, mutations in the heat-shock protein 90, which chaperones the maturation of many regulatory proteins, have been shown to uncover genetic variation that exists in a silent state in *Drosophila* and *Arabidopsis*. Thus, Hsp90 is thought to buffer morphogenetic pathways from the destabilizing effects of stochastic processes, allowing these pathways to accumulate genetic variants under neutral conditions. This is an example of canalization, that is, the relative phenotypic insensitivity of development to mutation or environmental changes (Waddington 1942; Rutherford and Lindquist 1998; Queitsch *et al* 2002). Cooperation at the population level involving feedback loops, as in the case of the community effect mentioned above, might also increase resilience in the response to genetic and environmental changes. The same is true for the existence of developmental checkpoints to make sure that one process is successfully finished before engaging in the next one in a developmental cascade. After that, one is tempted to think that noise is always bad. But is that really so?

One can question what the benefit of determining a whole tissue or organ from a limited cell population would be. In small organisms with short developmental times, it is obvious that primordial cell populations are limited. They are expected to be particularly sensitive to asymmetry. In larger organisms, with long developmental times, this is less obvious. However, one can argue that stochasticity provides a source of variability that is not encoded by the genome. Even in the case of clonal organisms, stochastic choices might operate on many monoallelically expressed genes, which would lead to a great deal of potential variability from one individual to an-

other. A straightforward testable hypothesis would be that clonal individuals (that pass through a sexual phase) are likely to determine their tissues and organs from very small cell populations, as a developmental strategy. It has also been proposed that during development of differentiated tissue from stem cells, as for blood, there must always be a stem cell subpopulation not expressing the factor that commits to the developmental pathway. This would leave enough primordial cells to ensure renewal of the tissue (Becskei *et al* 2001). If variation is the spice of life (Kruglyak and Nickerson 2001), then stochastic processes make life spicy, more than one can suspect by simply reading the letters in the genome. The same argument holds for the various examples of binary expression that have been mentioned above. From this perspective, multicellular organisms are thought of as mosaics of cells, each one expressing a subset of the genes characteristic of the tissue they form considered as a whole. In this view, each cell is singular and each organism is singular as well.

6. Conclusions

The above discussion leads us to consider the clones as ersatz copies rather than perfect replicants, because many stochastic events are interposed between the information encoded in DNA and its expression at the phenotypic level. Therefore, it is not surprising that most human monozygotic twin pairs may bear major discordances for birth weight, expression of genetic diseases and congenital anomalies (Machin 1996). They are not identical but simply monozygotic. This is true for handedness, as there is no correlation of left- or right-handedness between monozygotic twins (Corballis 1991). More tellingly, concerning X-linked diseases, there are dozens of discordant monozygotic female twin cases. In such twins, the X bearing the mutant allele is inactivated in most cells of the normal twin while the X bearing the normal allele inactivated in most cells of the affected one. Otherwise, there may be skewed X-inactivation in one twin and random in the co-twin (Tiberio 1994). One tends to believe that monozygotic twins are at least identical from the point of view of the tissue antigens. Although this is in principle so, there are cases of discordance with regard to autoimmunity (Gregersen 1993). For instance, in the context of systemic lupus erythematosus in childhood, there may be concordance for the presence of anti-DNA antibodies but discordance for disease expression (Bustabad *et al* 1997).

There are many recent examples of cloned mammals (extensive information at http://www.roslin.ac.uk/public/cloning.html). The prospect of human cloning is especially controversial raising ethical and philosophical concerns. Indeed, it is difficult to conceive the existence of our

own, perhaps younger, *alter ego* somewhere. But even if this ersatz *alter ego* were there, one has the right to think that he or she (it?) is basically different, for the many reasons we have adduced. In 1999, Erwin Chargaff asked whether the soul could be cloned. The answer to this question cannot be provided here. But, what is pretty clear from our line of reasoning is that the material entity that somehow bears the soul is unclonable. The amazing complexity of brain circuitry, where just a rough scheme is genetically pre-defined, makes this all the more certain.

References

Abkowitz J L, Taboada M, Shelton G H, Catlin S N, Guttorp P and Kiklevich J V 1998 An X chromosome gene regulates hematopoietic stem cell kinetics; *Proc. Natl. Acad. Sci. USA* **95** 3862–3866

Ayoub N, Goldshmidt I and Cohen A 1999 Position effect variegation at the mating-type locus of fission yeast: a cis-acting element inhibits covariegated expression of genes in the silent and expressed domains; *Genetics* **152** 495–508

Becskei A, Seraphin B and Serrano L 2001 Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion; *EMBO J.* **20** 2528–2535

Biggar S R and Crabtree G R 2001 Cell signaling can direct either binary or graded transcriptional responses; *EMBO J.* **20** 3167–3176

Blake W J, Ka M, Cantor C R and Collins J J 2003 Noise in eukaryotic gene expression; *Nature (London)* **422** 633–637

Blattner F R, Plunkett G 3rd, Bloch C A, Perna N T, Burland V, Riley M, Collado-Vides J, Glasner J D, Rode C K, Mayhew G F, Gregor J, Davis N W, Kirkpatrick H A, Goeden M A, Rose D J, Mau B and Shao Y 1997 The complete genome sequence of *Escherichia coli* K-12; *Science* **277** 1453–1474

Boulanger P A, Yoshinaga S K and Berk A J 1987 DNA-binding properties and characterization of human transcription factor TFIIIC2; *J. Biol. Chem.* **262** 15098–15105

Brown S W and Chandra H S 1973 Inactivation system of the mammalian X chromosome; *Proc. Natl. Acad. Sci. USA* **70** 195–199

Bustabad S, Gonzalez T and Trujillo E 1997 Systemic lupus erythematosus in childhood. Concordance in a pair of monozygotic twins for anti-dsDNA antibodies and discordance for disease expression; *J. Rheumatol.* **24** 1450–1451

Chargaff E 2000 Molecular slavery; *J. Biosci.* **25** 7–8

Cook D L, Gerber A N and Tapscott S J 1998 Modeling stochastic gene expression: implications for haploinsufficiency; *Proc. Natl. Acad. Sci. USA* **95** 15641–15646

Corballis M C 1991 *The lopsided ape* (Oxford: Oxford University Press), p. 97

Davey H M and Kell D B 1996 Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses; *Microbiol. Rev.* **60** 641–696

Dingemanse M A, de Boer P A, Moorman A F, Charles R and Lamers W H 1994 The expression of liver-specific genes within rat embryonic hepatocytes is a discontinuous process; *Differentiation* **56** 153–162

Elowitz M B, Levine A J, Siggia E D and Swain P S 2002 Stochastic gene expression in a single cell; *Science* **297** 1183–1186

Ehrlich J, Sankoff D and Nadeau J H 1997 Synteny conserva-

tion and chromosome rearrangements during mammalian evolution; *Genetics* **147** 289–296

Fiering S, Whitelaw E and Martin D I 2000 To be or not to be active: the stochastic nature of enhancer action; *Bioessays* **22** 381–387

Francastel C, Walters M C, Groudine M and Martin D I 1999 A functional enhancer suppresses silencing of a transgene and prevents its localization close to centrometric heterochromatin; *Cell* **99** 259–269

Gardner T S, Cantor C R and Collins J J 2000 Construction of a genetic toggle switch in *Escherichia coli. Nature (London)* **403** 339–342

Gregersen P K 1993 Discordance for autoimmunity in monozygotic twins. Are "identical" twins really identical?; *Arthritis Rheum.* **36** 1185–1192

Grunstein M 1998 Yeast Heterochromatin: Regulation of Its Assembly and Inheritance by Histones; *Cell* **93** 325–328

Gurdon J B, Kato K and Lemaire P 1993 The community effect, dorsalization and mesoderm induction; *Curr. Opin. Genet. Dev.* **3** 662–667

Gu Z, Steinmetz L M, Gu X, Scharfe C, Davis R W and Li W H 2003 Role of duplicate genes in genetic robustness against null mutations; *Nature (London)* **421** 63–66

Kato H, Nagamine M, Kominami R and Muramatsu M 1986 Formation of the transcription initiation complex on mammalian rDNA; *Mol. Cell. Biol.* **6** 3418–3427

Kellner J R and Alford R A 2003 The ontogeny of fluctuating asymmetry; *Am. Nat.* **161** 931–947

Klingenberg C P 2003 A developmental perspective on developmental instability theory, models and mechanisms; in *Developmental instability causes and consequences* (ed.) M Polak (New York: Oxford University Press) pp 14–34

Kruglyak L and Nickerson D A 2001 Variation is the spice of life; *Nat. Genet.* **27** 234–236

Lewis J, Slack J M W and Wolpert L 1977 Thresholds in development; *J. Theor. Biol.* **65** 579–590

Liyanage M, Coleman A, du Manoir S, Veldman T, McCormack S, Dickson R B, Barlow C, Wynshaw-Boris A, Janz S, Wienberg J, Ferguson-Smith MA, Schrock E and Ried T 1996 Multicolour spectral karyotyping of mouse chromosomes; *Nat. Genet.* **14** 312–315

Lyon M F 1962 Sex chromatin and gene action in the mammalian X-chromosome; *Am. J. Hum. Genet.* **14** 135–148

Machin G A 1996 Some causes of genotypic and phenotypic discordance in monozygotic twin pairs; *Am. J. Med. Genet.* **61** 216–228

McAdams H H and Arkin A 1997 Stochastic mechanisms in gene expression; *Proc. Natl. Acad. Sci. USA* **94** 814–819

Muller H J 1930 Types of visible variations induced by X-rays in *Drosophila*; *J. Genet.* **22** 299–334

Murray J D 1989 *Mathematical biology* (New York: Springer-Verlag)

Novick A and Weiner M 1957 Enzyme induction as an all-or-none phenomenon; *Proc. Natl. Acad. Sci. USA* **43** 553–566

Newlands S, Levitt L K, Robinson C S, Karpf A B, Hodgson V R, Wade R P and Hardeman E C 1998 Transcription occurs in pulses in muscle fibers; *Genes Dev.* **12** 2748–2758

Nijhout H F 2005 Stochastic Gene expression: Dominance, thresholds, and boundaries; in *The biology of genetic dominance* Veitia R (ed.) (Georgetown: Landes Biosciences) (in press)

Nutt S L and Busslinger M 1999 Monoallelic expression of Pax5: a paradigm for the haploinsufficiency of mammalian Pax genes; *Biol. Chem.* **380** 601–611

Ogawa Y, Lee J T and Xite 2003 X-inactivation intergenic transcription elements that regulate the probability of choice; *Mol. Cell.* **11** 731–733

Osborne C S, Chakalova L, Brown K E, Carter D, Horton A, Debrand E, Goyenechea B, Mitchell J A, Lopes S, Reik W and Fraser P 2004 Active genes dynamically colocalize to shared sites of ongoing transcription; *Nat. Genet.* **36** 1065–1071

Ozbudak E M, Thattai M, Kurtser I, Grossman A D and van Oudenaarden A 2002 Regulation of noise in the expression of a single gene; *Nat. Genet.* **31** 69–73

Queitsch C, Sangster T A and Lindquist S 2002 Hsp90 as a capacitor of phenotypic variation; *Nature* (*London*) **417** 618–624

Rutherford S L and Lindquist S 1998 Hsp90 as a capacitor for morphological evolution; *Nature* (*London*) **396** 336–342

Schweidler E V 1906 Uber schwankungen der radioaktiven umwandlung; *Comptes Rendus du Premièr Congrès International pour L'étude de la Radiologie et de l'Ionisation, Liège*, Sep. 12–14, 1905. H Dunod and E Pinat, Paris, Section de physique

Shang Y, Hu X, DiRenzo J, Lazar M A and Brown M 2000 Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription; *Cell* **103** 843–852

Siegele D A and Hu J C 1997 Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations; *Proc. Natl. Acad. Sci. USA* **94** 8168–8172

Sutherland H G, Martin D I and Whitelaw E 1997 A globin enhancer acts by increasing the proportion of erythrocytes expressing a linked transgene; *Mol. Cell. Biol.* **17** 1607–1614

Tiberio G 1994 M Z female twins discordant for X-linked diseases: a review; *Acta. Genet. Med. Gemellol* (*Roma*) **43** 207–214

Veitia R A 2003 A sigmoidal transcriptional response: cooperativity, synergy and dosage effects; *Biol. Rev. Camb. Philos. Soc.* **78** 149–170

Waddington C H 1942 Canalization of development and the inheritance of acquired characters; *Nature* (*London*) **150** 563–565

Wakimoto B T 1998 Beyond the nucleosome: epigenetic aspects of position-effect variegation in *Drosophila*; *Cell* **93** 321–324

Weinstein L S 2001 The role of tissue-specific imprinting as a source of phenotypic heterogeneity in human disease; *Biol. Psychiatry* **50** 927–931

Wijgerde M, Grosveld F and Fraser P 1995 Transcription complex stability and chromatin dynamics in vivo; *Nature* (*London*) **377** 209–213

Willard H F 1996 X chromosome inactivation, XIST, and pursuit of the X-inactivation center; *cell* **86** 5–7