Available online at www.sciencedirect.com
ScienceDirect
Gene 390 (2007) 232–242
GENE
SECTION
FUNCTIONAL GENOMICS
www.elsevier.com/locate/gene

Review

Combinatorial epigenetics, "junk DNA", and the evolution of complex organisms

Emile Zuckerkandl [a,*], Giacomo Cavalli [b]

[a] *Department of Biological Sciences, Stanford University, Stanford, California 94305; and Institute of Molecular Medical Sciences, P.O. Box 20452, Stanford, CA 94309, USA*
[b] *Institute of Human Genetics, CNRS. 141 rue de la Cardonille, F-34396 Montpellier Cedex5, France*

Available online 9 December 2006
Received by M. Batzer

Abstract

At certain evolutionary junctures, two or more mutations participating in the build-up of a new complex function may be required to become available simultaneously in the same individuals. How could this happen in higher organisms whose populations are small compared to those of microbes, and in which chances of combined nearly simultaneous highly specific favorable mutations are correspondingly low? The question can in principle be answered for regulatory evolution, one of the basic processes of evolutionary change. A combined resetting of transcription rates in several genes could occur in the same individual. It is proposed that, in eukaryotes, changes in epigenetic trends and epigenetically transforming encounters between alternative chromatin structures could arise frequently enough so as to render probable particular conjunctions of changed transcription rates. Such conjunctions could involve mutational changes with low specificity requirements in gene-associated regions of non-protein-coding sequences. The effects of such mutations, notably when they determine the use of histone variants and covalent modifications of histones, can be among those that migrate along chromatin. Changes in chromatin structure are often cellularly inheritable over at least a limited number of generations of cells, and of individuals when the germ line is involved. SINEs and LINEs, which have been considered "junk DNA", are among the repeat sequences that would appear liable to have teleregulatory effects on the function of a nearby promoter, through changes in their numbers and distribution. There may also be present preexisting unstably inheritable epigenetic trends leading to cellular variegation, trends endemic in a cell population based on DNA sequences previously established in the neighborhood. Either way, epigenetically conditioned teleregulatory trends may display only limited penetrance. The imposition at a distance of new chromatin structures with regulatory impact can occur in *cis* as well as in *trans*, and is examined as intrachromosomally spreading teleregulation and interchromosomal "gene kissing". The chances for two or more particular epigenetically determined regulatory trends to occur together in a cell are increased thanks to the proposed low specificity requirements for most of the pertinent sequence changes in intergenic and intronic DNA or in the distribution of middle repetitive sequences that have teleregulatory impact. Inheritable epigenetic changes ("epimutations") with effects at a distance would then perdure over the number of generations required for "assimilation" of the several regulatory novelties through the occurrence and selection, gene by gene, of specific classical mutations. These mutations would have effects similar to the epigenetic effects, yet would provide stability and penetrance. The described epigenetic/genetic partnership may well at times have opened the way toward certain complex new functions. Thus, the presence of "junk DNA", through co-determining the (higher or lower) order and the variants of chromatin structure with regulatory effects at a distance, might make an important contribution to the evolution of complex organisms.

Keywords: Intron functions; GC-rich isochores

Evolutionary history attests that the most complex ("highest") organisms are still evolvable. The question is, how. These organisms are considerably larger than the lowest organisms. As a consequence, the highest organisms also have in common the property of producing much smaller populations. They must thus be assumed to possess a more limited store of pre-established mutations required for adaptive responses and generate a low expectation for simultaneously adaptive mutations that are yet to occur. It thus seems particularly unlikely in the case of the highest organisms that more than one

* Corresponding author.
E-mail addresses: emilez@stanford.edu (E. Zuckerkandl), giacomo.cavalli@igh.cnrs.fr (G. Cavalli).

0378-1119/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.12.001

nonrecessive and adaptive mutation relating to the same evolving function occurs in the same germ cell at nearly the same time.

On the other hand, in regard to molecular lock-and-key interactions, it often may at first seem awkward to postulate the successive independent evolution of the key and of the lock (Adami, 2006). More generally, at the morphological level, many evolutionary developments are based on multiple co-adaptations that only make sense in terms of their simultaneous presence (Cuénot, 1941). Such co-adaptations are essentially of a regulatory nature. Darwin (1872) knew that it was not always easy to explain simultaneous presence short of simultaneous occurrence. Yet, he correctly surmised that complex biological systems are indeed explainable in terms of "successive slight modifications".[1] This expectation has been strongly supported by the discovery of multiple simultaneous functions and functional potentials in macromolecular components and of their divergence after gene duplication. Such a system provides a store of resources for new sets of gene interactions and for the use of old interactions in new combinations. Darwin's contention regarding pathways for the stepwise building of biological systems has recently been verified *in silico* by Lenski et al. (2003) and has now been demonstrated in an actual, typically challenging case of molecular evolution (Bridgham et al., 2006). The authors showed how diversifying key-and-lock partners co-adapted by stepwise changes. Evolution's boundless addiction to tinkering permits it to resort to successive modifications even when the existence of such a pathway is at first not obvious to the observer.

Single-step solutions to organizational problems do not imply, however, that the simultaneous or near-simultaneous occurrence of two or more mutually co-adapted mutations would not facilitate certain evolutionary processes and perhaps open the door to others. It is therefore of interest to point out that in regard to transcriptional regulation conditions may well be right, even in the rather small effective populations of the most complex eukaryotes, for the simultaneous occurrence in single individuals of mutually co-adaptive inheritable changes. The present paper describes a process that supports this assertion, which, if correct, could widen the spectrum of biological dispositions favorable to the evolution of the highest organisms.

Some such dispositions are genetic (i.e., dependent on DNA sequence) and some epigenetic (dependent on the structure of chromatin), though the epigenetic chromatin structure also depends on DNA sequence, when not directly, then indirectly and partially. Epigenetics is widely considered as relating to a body of processes whereby gene expression is modified inheritably by mechanisms that do not involve DNA mutation. During development the difference between genetic and epigenetic processes is quite clear-cut, since in development chromatin structure changes locally or regionally, whereas DNA sequence structure mostly does not. In evolution, on the other hand, contributions of genetic and epigenetic change can occur in different DNA regions and their effects can be combined. In that context, both genetic and epigenetic processes of regulatory change may well frequently make use of various sorts of so-called junk DNA, notably of one very common form of "junk", namely, varieties of retrotransposons called SINEs (Short Interspersed Elements) and LINEs (Long Interspersed Elements). Through the retrotransposition of such elements, genetic kinds of effects seem to be largely obtained by the distribution over mammalian genomes of a heretofore unimagined wealth of additional factor-binding DNA motifs (Polak and Domany, 2006; Polak, 2006). Epigenetically, on the other hand, retrotransposons presumably often function as co-determinants of the order of chromatin structure, higher or lower.

We ask today a question that many no doubt would still judge rather extravagant: is not a certain amount of so-called junk DNA instrumental and even pivotal for some of those relatively rare steps leading to the generation (especially) of the most complex organisms? In order for DNA so characterized to play such a role, the conjunction of a set of macromolecular structures and processes would be required. Many of these have been individually established or inferred, but they may not have been considered collectively, and certain implications of and connections among the processes may therefore not have been brought into focus. These implications and connections are linked to a set of preconditions for any regulatory effects on gene transcription that would originate in "junk DNA". There is much evidence for genomic mechanisms fulfilling these conditions.

1. Some conditions under which "junk DNA" may help generate higher organisms

Genomes need to possess at least the following six basic properties (Table 1) if any sequences considered as "junk

Table 1
Some preconditions under which "junk DNA" can help generate higher organisms

1 – A distinction between genetic and epigenetic effects applies
2 – Quantitative changes in gene expression and changes in gene interaction topology that at times directly ensue play eminent evolutionary roles
3 – Local changes in chromatin can have effects on chromatin structure at a distance
4 – A variety of different local genetic and epigenetic characteristics can have similar regulatory effects at a distance
5 – Certain epigenetic modifications are inheritable
6 – Epigenetic effects at a distance tend to be all-or-none. Regulatory adjustments in the transcription rate of a gene can rely initially on the dosing of cellular variegation

[1] The possibility that one-step developmental modifications include at times a large single step need not be excluded. If large-step modifications do (rarely) occur, they are however quite unlikely to lead to any significant upward jump in complexity, provided redundant genes and gene interactions are excluded from complexity measurements. Complexity no doubt increases only by small steps during evolution. This statement is not contradicted by polyploidy whose great contributions to complexity, when made, can only be realized over time. On the other hand, biological complexity can probably *decrease* by larger single steps, notably in the case of parasitism. Biological complexity can be in part described as interaction complexity involving informational macromolecules, but is very difficult to define in a way that is exhaustive and yet not redundant. No attempt at defining it is made here.

DNA" are to intervene in organismic evolution, as directly linked to the evolution of genes (leaving aside chromosomal evolution).

First, the distinction between genetic and epigenetic effects must apply. The existence of epigenetic effects is not episodic; it is at the core of living systems. The roles in chromatin of the polynucleotides and their interacting factors are inseparable and *equally causal*, a circumstance not sufficiently emphasized by those who consider that DNA by itself represents "the script of life" (Zuckerkandl, 2002). If junk DNA is to be included as a partner in functional evolution, it can only be through the interactions of *factors* with this DNA.

Epigenetic effects on gene expression, as indicated, are now usually conceived as changes in gene expression not linked to changes in DNA sequence. For the present purpose, we let the term epigenetic refer to changes in gene expression that are brought about by changes in chromatin, without specifying whether or not *somewhere* in the DNA, at a distance, a sequence change has occurred that led to the regulatory effect in the gene considered. The condition for characterizing some regulatory effects as epigenetic then is only that any causally linked sequence change, if present, did *not* take place in the gene's primary regulatory dependencies. By primary regulatory dependencies one may designate the relatively short and specific regulatory sequences such as promoters, enhancers, and insulators. Inheritable epigenetic changes may be called epimutations, a term first used by Waddington (1942). The penetrance of epimutations may remain limiting (see below) and their stability limited. It should be kept in mind that *inheritable* epigenetic changes likely correlate with features of DNA sequence, even though these features may be far from uniquely defined.

The second precondition for attributing to mutational events in junk DNA a role in the functioning of neighboring genes is that the evolution of gene *regulation* plays a pivotal role in the evolution of organisms. That it likely does has been claimed about four decades ago (Zuckerkandl, 1963, 1964, 1968, 1983; Zuckerkandl and Pauling, 1965) and has become increasingly well established (King and Wilson, 1975; Li and Noll, 1994). Most often critical evolutionary changes may not consist in introducing new cellular functions by modifying effector proteins (proteins other than regulatory factors). Instead, innovation takes the form of novel combinations in the quantitative and spatiotemporal regulation of preexisting informational macromolecules. Regulatory changes are in fact the only ones that junk DNA can bring about, barring the occasional use as exons of some sequence fragments (Makalowki et al., 1994; Nekrutenko and Li, 2001). (Another area of control by "junk DNA" is the regulation of chromosome behavior.) Changes in the topology of regulatory networks are introduced by several types of events, but even quantitative changes alone in gene transcription are capable of abolishing some old interactions and of establishing new ones. The accidental mutual tuning of several simultaneous epigenetic modifications within the framework of a newly deployed part of the gene interaction network is presumably sufficiently probable, since the number of effectively different combinations of regulatory changes in two or even three genes may mostly not be very large.

The range of variability in the expression of individual genes in populations of cells or of organisms is now known to be far in excess of what had been expected (cf. Rubin, 1990; Brem et al., 2002; Cowles et al., 2002; Pastinen et al., 2003; Lo et al., 2003; Bray et al., 2003; Cheung et al., 2003; Knight, 2004). Such variability is found even in genetically identical cells and organisms that have identical histories of environmental exposure (Raser and O'Shea, 2005). An important part of these inequalities in gene expression represents variations in the amounts of transcription products. Collectively, the inequalities of gene expression in "identical" cells have been referred to as "noise" (Raser and O'Shea, 2005), but some are attributable to controlled periodic fluctuations in cellular components (Rosenfeld et al., 2005). An as yet undefined fraction of this noise is due to inheritable epigenetic allelic variation. It turns out that diploid genomes, at least in higher organisms, are constituted not only of sequence haplotypes, but also of regulatory haplotypes. One manifestation thereof is differential allelic DNA methylation (Silva and White, 1988). The prospects are good for being able to attribute to epimutations a fraction of the observed variabilities in gene expression, and this fraction can be expected to prove significant in terms of absolute numbers of occurrences per cell.

There is an obvious third precondition, to be detailed later, for gene-associated "junk DNA" to play a distinctive and perhaps widespread role in regulatory evolution. It would be required that certain local changes in chromatin structure can have structural repercussions *at a distance*. Certain chromatin structures appearing within "junk DNA" indeed tend to spread *in cis*, and when they do, they may be expected often to spread toward focal sequence elements of transcriptional control such as promoters or enhancers. Local structural conformations also tend to spread in *trans* under certain conditions, as will be discussed. To be effective, processes in *trans* in turn no doubt require that the modifications in chromatin structure of the target sector encompass a gene's primary regulatory dependencies, notably promoters and/or enhancers.

A fourth precondition of the present model of action of junk DNA in gene regulation is applicable to structural spreading in *cis*. Not only must the particular location of a teleregulatory mutation be permissive but the nature of the mutational change must be permissive, too. In order for the process envisaged to be of significant frequency, a *variety* of mutational changes in junk DNA must indeed be capable of resulting in similar regulatory effects at a distance. Various links in the chains of regulation-altering processes can be at the origin of the events. DNA methylation and histone modifications, for example, are linked at least indirectly. The binding of short interfering RNAs (siRNAs) to promoters induces DNA methylation and is reversed by the inhibitor of DNA methylation 5-azacytidine and also by TSA, an inhibitor of histone deacetylases. This suggests that DNA methylation is linked to histone deacetylation in this case (Morris et al., 2004). In *Arabidopsis*, the majority of methylated sequences are repeat sequences, and most repeats are derived from transposable elements (Rangwala

and Richards, 2004). Thus many transcription-altering mutations occurring at a distance could consist in insertions or deletions of repeat sequences such as SINEs or LINEs.

As a fifth precondition for the use of "junk DNA" in support of the evolution of transcriptional regulation in complex organisms, certain changes in chromatin structure must be *inheritable* over at least a limited number of generations. Indeed, repressive complexes containing chromodomain proteins such as Polycomb or HP1 or their homologues are mitotically as well as meiotically inheritable (Cavalli and Paro, 1998). Epigenetic inheritance is found in particular in sectorially controlled developmental genes such as the Hox genes, analyzed notably by the groups of Renato Paro, Giacomo Cavalli, and Vincenzo Pirrotta (see Zuckerkandl, 1999). The *Drosophila* developmental regulatory gene, *wingless*, very likely also is sectorially controlled (Tolhuis et al., 2006). In a category of genes and especially of genes grouped in complexes, sectorial gene control is present as a mode of transcriptional regulation in addition to the universal "punctate" regulation, which is exercised through a gene's shorter specific primary regulatory dependencies such as promoters and enhancers. In sectorial regulation, the coding sequences and their primary regulatory dependencies can be either sequestered or maintained in an "open" structure thanks to alternative chromatin conformations extending over a DNA sector that is at least tens of kilobases long (Zuckerkandl, 1997). The *wingless* gene can be considered to be one of the developmental regulatory genes whose chromatin sector, thanks to the presence of Polycomb group response elements, trithorax group response elements, and their corresponding factors, is subject to switching between inheritable transcriptionally potentiated and "superrepressed" (sectorially repressed) states, according to developmental circumstances. The switch to the superrepressed state of the *wingless* gene can be obtained by various functionally damaging mutations in coding sequences for trithorax group factors (Sollars et al., 2003). The same state can also be obtained through a mere reduction in factor dosage without any sequence change in the DNA, as expected—an epigenetic change such as occurs routinely during development. Whether the switch moves as a consequence of a mutation or merely of an epimutation, Sollars et al. find that it stays put in its new position without the mechanism of switch turning having to intervene further, again as expected in the presence of built-in stability of the bimodal switch positions in sectorially controlled genes. An epigenetic regional linearly progressive restructuring of chromatin or transfer in *trans* of a chromatin structure, can at times represent intermediate evolutionary stages, on the way towards the genetic "assimilation" through ordinary mutations of an epigenetic regulatory change.

Structural states of chromatin required for *transcriptional activity* are inheritable through a nucleosome replacement process linked to transcription (cf. Henikoff et al., 2004a). Small interfering RNAs (siRNAs) play central roles in establishing and maintaining inheritable states of gene activity, notably in the plant *Arabidopsis* (Zilberman and Henikoff, 2005). Inheritable transcriptional potentiation is accompanied by H3K4 methylation and other specific histone modifications (Kouskouti and Talianidis, 2005). On the other hand, *transcriptional silencing* of the epiallele of a transgene correlates in mammals with its DNA methylation, even though epigenetic inheritance of repression is found also in flies and yeast which lack DNA methylation (Richards and Elgin, 2002; Henikoff et al., 2004a,b). More generally important regarding the establishment of an inheritable repressive state may be the secondary histone modifications, such as in the case of a switch to the repressed state of chromatin, deacetylation of H3K9, demethylation of H3K4, methylation of H3K9, and methylation of cytosine in CpG dinucleotides. In the silencing of the promoter of the gene for murine terminal transferase (Dntt), the modifications occur in that order when immature thymocytes are stimulated to differentiate (Su et al., 2005), and the resulting chromatin structure is inheritable. Secondary histone modifications may be a general condition for the regional formation of repressive chromatin, whereas the addition to the system of PcG factors (Cao et al., 2002; Tolhuis et al., 2006) characterizes the particularly stably repressed sectorially regulated genes. Cellular inheritability, which thus includes the switch positions of such genes, apparently applies beyond this system to all repression correlating with the presence of certain histone variants and histone secondary modifications. We suggest that, through such modulations in histones among other structural determinants, the results of teleregulation can be inheritable in *cis* as well as in *trans* and as repression as well as transcriptional potentiation of genes.

A sixth precondition, or at least favorable circumstance, may well be that the epigenetic effects at a distance lead to the gene's variegation. It seems indeed rather probable that a regulatory change in the transcription of a neighboring gene through the type of mechanism envisaged will be "all-or-none" rather than display intermediate values, because intermediate values would require stable intermediate structures, and there presumably are no stable intermediate structures along the pathway of a binary switch. Indeed, for example, Sutherland et al. (2000), at the level of individual cells, found all-or-none effects of silencing epimutations. Often a new mutational event need not occur in order for an epigenetic trend to be manifest and exploited. The presence of a "trend" implies that the variant chromatin structure is expressed in a minority of cells within a cell population. Yet, it may well be fully expressed when expressed at all. On the other hand, a rate change in the gene's transcription may often be adaptive when moderate, but not when radical as would be the case of stable outright silencing of a gene in all cells of a cell population. Structural effects in *cis* of SINEs, especially locally grouped SINEs, or of other sequence repeats, may precisely give rise, in many cases, to repressive high-order structures in the chromatin neighborhood (see below). It may thus often be of selective advantage to the organism that the penetrance of new high-order structures induced by transposons be limited—as they may well tend to be rather generally. This limitation would imply variegation, namely, the structures would be formed in some cells and not in others. Even though at the level of individual cells an effect of "opening" or "closing" chromatin would be all-or-none, at the level of a tissue the production level would be intermediate and

could be desirably moderate. Implicated here would be proteins engaged on the pathways of endocrine and exocrine activity and proteins forming developmental gradients specifically engaged in the gradient-forming proteins' manufacture and delivery. Variegation could thus at times have a biological function, although, when a classical mutation eventually stabilizes the epimutation or replaces it, only a temporary one.

Such variegation is expected either to be introduced de novo in the neighborhood of a gene through some DNA sequence change or, if slight variegation preexisted in a more or less cryptic way in the same chromatin region, it could be intensified through selection, in the sense of a new structural state of chromatin spreading to a progressively larger proportion of cells in a given cell population. One may refer to an intermediate structural provision on its way from the original transcriptional regulation of a gene in some cells to a change in that regulation generalized over a population as a *variegation bridge* spanning the distance between a partially penetrant epigenetic opportunity turned adaptive and its genetic assimilation leading to full penetrance of an appropriate regulatory state. In the case of linear spreading along the chromosome, the structural change may well not progress equally far in individual cells, and the epigenetic effect on transcriptional regulation in a neighboring gene, again, may therefore be variegated. Structural spreading in *trans*, i.e., interchromosomal structural transmission, may be variegated as well.

As discussed or to be discussed, the six preliminary conditions enumerated and listed in Table 1 are in fact all fulfilled—some of them at least under certain circumstances. Therefore, higher eukaryotes may well be able to play with novel combinations of gene activities without having to wait for highly specific mutations to permit them to do so.

2. Degrees of specificity of regulatory effects originating in "junk DNA"

A general way for "junk DNA" to intervene in gene regulation no doubt depends on the capacity of this DNA to affect the balance of bound and free regulatory factors. Transcription factors extensively bind to non-protein-coding sequences (Cawley et al., 2004), either in functionally neutral associations, though often leading to perhaps functionally nonneutral local transcription of noncoding RNAs, or with effects on genome structure endowed with functional implications. In particular, factor binding may have teleregulatory effects (see below). Factor binding can lead to the formation of local high-order structures of chromatin, notably based on interacting sequence repeats (Dorer and Henikoff, 1994; Sabl and Henikoff, 1996; Hsieh and Fire, 2000; Trifonov, 2004). The relationship between sequence repeats in euchromatin and in heterochromatin (which in turn contains simple-sequence repeats) no doubt depends in part on the epigenetic state of the repeats in euchromatin, e.g., on their methylation or nonmethylation. High-order structures formed by sequence repeats in euchromatin can be stabilized by their interaction with an accessible zone of constitutive heterochromatin (Dorer and Henikoff, 1997). On the other hand, because heterochromatin presents factor binding sites (e.g., Torok et al., 2000), factors participating in high-order structure formation by euchromatic sequence repeats can in principle be titrated down by increases in the amount of available heterochromatin (Zuckerkandl, 1974; Locke et al., 1988). The inference is that heterochromatin should be able to promote as well as inhibit the heterochromatinization of sectors of euchromatin. Similarly, a regional "melting" of heterochromatin might be expected to lead to a release of factors that would intensify a looming heterochromatinizing effects on lower-order chromatin structures. Any shifting of the equilibrium between heterochromatin and euchromatin in the nucleus is thus liable to have effects on structures in euchromatin. In the presence of a particular combination of factor activities and according to cell type, regions primarily implicated would be the ones that are close to a flip-over point of their chromatin structure. Genomes thus may function as factor titrating machines that would operate through developmental as well as evolutionary changes in the amounts of heterochromatin-like material in the nucleus. Large amounts of non-protein-coding sequences in large genomes thus are expected to play functional roles, provided the regulatory factors to be bound or released are not present in great excess. Dissociation constants, which relate to factor concentrations at which one half of the factors are bound within their specific types of complexes, are probably not higher, in general, than the nuclear concentrations in these factors. Hence, factor concentration is indeed a sensitive limiting variable in the efficiency of binding (cf. Chambeyron and Bickmore, 2004), and a large number of accessible nonspecific binding sites and variations in this number are likely to have their regulatory role to play.

A somewhat more specific mode of gene regulation implicating "junk DNA" concerns the control of chromatin structure in and around genes. It would seem rather obvious that sequences whose primary task is to code for proteins may likely be, on average, relatively inadequate when it comes to carrying out other competing tasks, notably when high-order chromatin structures are required to be stable under a variety of cellular conditions. It has therefore been proposed a long time ago (Zuckerkandl, 1981) that the function of introns ("junk") was to make up for this and other presumed deficiencies of exons (not "junk"). Introns thus would compensate for the shortcomings of exons, which because of special constraints dictated by protein function may well fail in some respects as citizens of genomes. In particular, tissue-specific genes probably require stability of sequestration in all the cell types in which these genes are repressed. The surmise that introns are better qualified than exons in this respect has been supported by the observation of higher nucleosome formation potentials in introns than in exons (Levitsky et al., 2001). Other functional traits that exons may not present to a sufficient degree are sequence conditions for maximizing transcription rates. Indeed, the GC content of introns correlates with the maximum level of gene expression among tissues, as confirmed by Vinogradov (2005). The implication seems to be that, at least in GC-rich isochores, the compositional characteristics of introns (even though they are of reduced length in these regions) favor high transcription rates

(Bernardi, 2001, 2004; Castillo-Davis et al., 2002). In fact, transcriptional levels have been found to be correlated with GC-rich isochores (D'Onofrio, 2002). The causal role in this correlation played by GC richness, if verified, would point to one of the functional effects of GC-rich isochores.

In high-GC, permanently "open" isochores, a trend toward structure building is probably counterproductive, and selection pressure in favor of compactness at the level of nucleotide pluralities (a compactness perhaps sought also for additional reasons) seems to prevail, in part through the limitation of intron length, though not generally to the extent of eliminating introns altogether. When introns impose their presence in high-GC regions, albeit a more modest one, it might be on account of the introns' compositional function. All the same, with intergenic sequences reduced as well, high-GC isochores can become very gene-dense.

On the other hand, those genes that under certain conditions are to be sequestered during interphase in chromatin structures of a high order – genes mostly located in low-GC areas of the human genome, as is the beta-globin gene complex – do not seem able to dispense with larger amounts of "junk DNA". Thus, both general functions envisaged here for introns, namely, ensuring the stability of high-order chromatin structures and achieving highest transcription rates in low-order chromatin structures, might be major causes for the *de novo* introduction of introns into protein coding sequences.

Too many unknowns still permeate some of the major regulatory contributions of "junk DNA". These contributions will not be considered despite their probable significance. A vast array of noncoding RNAs, part of which function as specific breakdown products obtained from larger transcripts, are derived from intergenic as well as genic regions. Such RNAs partake, often decisively, in the determination of chromatin structure and in the pretranscriptional as well as posttranscriptional control of gene expression. It is of little help that no satisfactory agreement has yet been reached, from a contemporary perspective, on what a gene and therefore an intergenic sequence exactly are (Pearson, 2006).

In addition to exercising general regulatory effects, "junk DNA" can also intervene in the regulation of specific genes. We wish to discuss here a process of this kind.

3. Junk DNA-dependent teleregulation of chromatin structure via spreading of chromatin modifications in *cis*

Regarding the question of the role that "junk DNA" located in the region of particular genes may play as originators of changes in transcriptional regulation, let us first consider effects in *cis*. We presume that, not infrequently, in some points of the genome, and within non-protein-coding DNA, a new chromatin structure is formed and is communicated to other points of the genome. The local nidus formation from which a new structure can spread or an old one can be selected for greater penetrance may take two categories of forms, activating or repressive.

In a process of local transcriptional potentiation, a moderately specific origin of transcription may be mutationally established or, when it is already present, the frequency in a cell population of the transcription of some intergenic or intronic sequence may be increased through the selection of sequence variants or variants of the cellular regulatory background. Such adventitious intergenic or intronic transcription might have either activating (Ling et al., 2004) or interfering (Allen et al., 2004) effects on the transcription of a protein coding sequence at a distance. Chromatin structures open to transcription may be able to be propagated linearly, and through transcription itself. The passage of RNA polymerase II probably can precede the transcription-friendly hyperacetylation of histones H3 and H4 as well as transcription-friendly hypermethylation of lysines 4 and 79 of histone H3, in both *Drosophila* and mouse (Ng et al., 2003; Schubeler et al., 2004; Kimura et al., 2004). When transcription extends into sectors where H3K4 has not been methylated, it may itself favor the generation of a stabilized open form of chromatin that H3K4 contributes to determine. Cellularly inheritable transcriptional activity of a sectorially repressible gene such as a *Hox* gene can be established in the early embryo and durably oppose repression (Poux et al., 2002). It has been shown in human hepatic genes that covalent secondary histone modifications such as H3K4, associated with active genes, migrate about 5 kb down the coding sequences, though it was not clear that this distance was a limit (Kouskouti and Talianidis, 2005). In the human beta-globin complex, the activation of the epsilon-globin gene requires the transcription – even though in this case in bits and pieces (Ling et al., 2005) – of the whole DNA sector from the distal HS2 enhancer to the epsilon-globin promoter (Ling et al., 2004), establishing the transcription of the intervening "junk" as a precondition for gene expression. Thus "junk RNA" can also be regulatory and often is, at least through the act of being transcribed (Schmitt et al., 2005) when not through the nature of its sequence.

The other category of gene-specific processes thought to intervene in *cis* consists in the formation of migrating *repressive* high-order chromatin structures, through secondary covalent modifications of histones and DNA. Methylated DNA, in particular, will adopt an inactive chromatin conformation (Antequera et al., 1989; Levine et al., 1991; Jaenisch and Bird, 2003). This inactive conformation can spread from a methylated to an adjacent nonmethylated region of DNA and thereby inhibit gene expression (Kass et al., 1993). The spreading of methylation from foci of methylated CpG sites seems to occur when DNA is rendered single-stranded at replication (Lindsay and Adams, 1996) and has been demonstrated in several systems (see Turker, 1999). In particular, the spreading of DNA methylation originating from human Alu elements and corresponding mouse B1 elements has been documented (cf. Yates et al., 1999; Jones and Takai, 2001), as well as from the plant SINE S1 (Arnaud et al., 2000). Both methylation of cytosine in CpG dinucleotides (Jaenisch and Bird, 2003) as well as methylation of certain lysine sites within histones (Jenuwein and Allis, 2001) are associated with transcriptional repression and are both capable of spreading along DNA. In the multiply interacting molecular modifications leading to repression, the methylation of lysine at position 9 of histone 3 ("H3K9") has been found to be causal

(Snowden et al., 2002). The mechanism of the structural "contagion" may not always be the same. Structural change can progress along chromatin or be communicated at a distance through loop formation, unless both processes occur at once (Kim and Dean, 2003; Zhao and Dean, 2004). In the case of histone methylation, the modification could also spread unevenly over a number of sites in a given segment of chromatin once that segment has been reached (Yates et al., 2003). The migration of modified structures along chromatin seems to be based on processive enzymatic activities in which the protein-binding domain of a factor recognizes the product of the factor's enzymatic domain and communicates the modification pattern to the next histone octamer (see Schreiber and Bernstein, 2002). The spreading over chromatin of secondary covalent modifications of histones has been observed repeatedly (e.g., Snowden et al., 2002), with a series of histone modifications spreading bidirectionally (Su et al., 2004; Su et al., 2005). Felsenfeld and his associates have shown that condensed chromatin structures tend to propagate themselves when they are not stopped by histone H3 and H4 acetylation and by lysine 4 methylation of histone H3 (West et al., 2004; Mutskov and Felsenfeld, 2004). A progressive spreading of repression may well be at the origin of the gradual extinction of the expression of a transgene during long-term propagation of cells in culture, an extinction not attributable to DNA methylation (Mutskov et al., 2002), and thus probably connected, by elimination, with histone modifications.

If a particular chromatin structure can be nucleated at various sites within non-protein-coding DNA, it is likely that the distance between such a nucleation zone and the transcription start site of a neighboring gene can be expected to vary rather widely according to circumstances. A broad range of nearly equivalent target sites for mutations should thus be potentially effective in teleregulation. Indeed, Kass et al. (1993) have shown that, in the spreading of methylated DNA, the extent of inhibition of a nearby gene is independent of the position of the methylated patch. Such flexibility in regard to site and precise molecular definition of the origins of a new structure leads to frequency of the effect, a frequency that confers value on combinatorial epigenetics.

Admittedly, in many genome locations the communication at a distance in *cis* of a structural feature of chromatin will be inhibited by insulators. The action of insulators, however, depends on the presence of certain factors (notably CTCF, Burgess-Beusse et al., 2002; Yusufzai et al., 2004) and is not invariant (Jeong and Pfeifer, 2004). The progression of methylation seems to take place after each round of DNA replication (Arnaud et al., 2000). When insulators do not radically reduce the expected relatively high frequency of the assumed processive structural change, or at other times are not present at all in the pathway of the structural procession, then mutually adaptive multiple regulatory changes can be expected occasionally to occur by accident. In this manner, certain changes in chromatin structure within "junk DNA" would create in individual cells a field of extensive natural experiments with various regulatory combinations.

These natural experiments are considerably extended by regulatory effects of chromatin in *trans*.

4. Teleregulation of chromatin structure via spreading of chromatin modifications in *trans*

Indeed, another class of phenomena that might lead to teleregulatory effects relies on "gene clustering" or "gene kissing" processes, whereby genomic elements sharing sequence homology or analogous molecular regulatory complementary structural features meet in the three-dimensional space of the cell nucleus. Upon mutation of one interacting partner (that we may designate as "donor" locus), a change in chromatin structure may thus be communicated to the "acceptor" locus in contact with the donor. In the case of gene activation, this type of structural and functional transfer has been documented only as represented by the phenomenon of transvection (for review see Duncan, 2002) and, to a lesser extent, in some recently observed cases of the phenomenon of gene kissing (Bantignies et al., 2003; Spilianakis et al., 2005; Ling et al., 2006). Compared to spreading in *cis*, this type of chromatin transmission might require an additional strong ability of components of the interacting partners to stabilize their contact via homo- or heterodimerization. This feature might be a basic property of heterochromatin, and, in this sense, "junk DNA" interactions leading to regulatory chromatin modifications between remote genomic loci might represent a special case of clustering of repeated DNA elements in the genome.

Gene kissing has been documented in *Drosophila* in the case of silencing mediated by heterochromatin (Dernburg et al., 1996; Csink and Henikoff, 1996; Harmon and Sedat, 2005), as well as by Polycomb group (PcG) proteins (Bantignies et al., 2003; Grimaud et al., 2006). Many other gene-kissing events have been reported in plants (Abranches et al., 2000). as well as in mammals, where they are not restricted to cases of heterochromatin-mediated silencing (Brown et al., 1997, 1999). They also occur in X-inactivation, a phenomenon of gene silencing that involves PcG proteins (Bacher et al., 2006; Xu et al., 2006), in genomic imprinting (Ling et al., 2006), and in activation of transcription (Osborne et al., 2004; Spilianakis et al., 2005; Brown et al., 2006). On a global genomic scale, human chromosomes frequently contact different chromosomes (Branco and Pombo, 2006). These contacts do not only affect gene expression but also enhance the likelihood of generating chromosome rearrangements, as illustrated by the fact that chromosomal translocations frequently occur between kissing loci (Nikiforova et al., 2000; Roix et al., 2003; Branco and Pombo, 2006). Importantly, these contacts may frequently involve repetitive sequences, although they are certainly not restricted to "junk DNA". In yeast, it has been recently shown that clustering of many loci involves sequences containing series of binding sites for the TFIIIC transcription factor (Noma et al., 2006). In *Drosophila*, heterochromatin repeats drive colocalization of the bw^D allele with pericentric heterochromatin (Dernburg et al., 1996; Csink and Henikoff, 1998; Harmon and Sedat, 2005). In mammals, heterochromatin regions cluster in cell nuclei (Ma et al., 2005). Moreover, most chromosome rearrangement events deriving from non-allelic homologous recombination are likely to be triggered by repetitive elements present at the two recombining (i.e. interacting) loci (reviewed

in Coghlan et al., 2005). Since LINEs and SINEs are present in many copies dispersed in the genome, it is likely that these elements are involved in gene contacts.

Gene kissing may cause teleregulation events using different mechanisms. First, acquisition of a new chromatin feature upon mutation or epimutation events, i.e. changes in histone modification marks or in the proteins recruited at the "donor" locus, may be followed by transmission to the acceptor loci of these new chromatin features via cooperative chromatin modification mechanisms. As an example, recruitment of a new histone modifying enzyme may modify histones also at loci in *trans*. Second, mutations or epimutations at the donor locus may change its interacting partners in the nucleus, by either loss of old interactors, acquisition of new interactors, or both. The chromatin state of the donor might thus be donated to a new set of remote "acceptor" loci. In the cases that have been studied, relatively small sequence elements of the order of several kilobases are sufficient to drive gene kissing. Thus, insertion of LINEs might be a sufficient event to change the three-dimensional nuclear environment of the donor locus, although larger rearrangements are likely to be as effective or even more.

It is of interest that *in vitro* not only histone methylation but also DNA methylation can be transferred in *trans* (Lindsay and Adams, 1996). It remains to be seen to what extent chromatin structure might stand in the way *in vivo* of such DNA/DNA interactions via a bridging methyltransferase molecule, but the potential for that type of kissing should not be ignored.

Finally, it must be stressed that the definition of donor and acceptor loci is an operational one. If one of the two loci is more tightly regulated and impermeable to changes in the nuclear environment, it might be more likely to behave as a donor and will not be susceptible to accept chromatin modifications in *trans*. In other cases, both loci may modify each other upon contact, and thus the separation of donor and acceptor may not always be clear-cut.

While teleregulation through linear structural diffusion of a structure along the chromosome can, in principle, alter not only quantitative features of gene expression but also the composition of gene interaction networks, structural transfers in *trans* potentially exert this second effect to a distinctly greater extent.

5. Wider implications of combinatorial epimutations

Selectable inheritable combinations of epigenetic changes in the transcription rate of individual genes can presumably be maintained in cells or organisms over a certain window of numbers of generations. That window can be of sufficient width to permit much rarer specific mutations with equivalent effects to occur and to be selected. These specific mutations would be of high penetrance and stability. They would produce the "assimilation" of certain constellations of epigenetic transmutations. This stabilizing and penetrance-promoting assimilation would be expected to occur gene by gene, whereas the innovative offer to the organism would have been made by two or several genes at a time. A new functional system could continue to be built in the process, with new epigenetic transmutations drawn into a growing novel gene interaction pattern based on preexisting genes or their slightly diverged duplicates. New assimilating genetic mutations would catch up with the epigenetically driven evolutionary progression. This genetic/epigenetic partnership and interplay in eukaryotes, already examined by Jablonka and Lamb (1995), would represent a pathway from a quick and provisional generation of relatively complex adaptive novelties to a more permanent state of the new regulatory pattern and one with in general full penetrance of the pattern. Importantly, the genetic/epigenetic partnership might well also, in fact, be the only way for organisms successfully to get across certain bottlenecks of functional evolution if the evolving systems did require several simultaneous or nearly simultaneous regulatory modifications.

In summary, if the proposed evolutionary mechanisms apply, they can bring together in germ cells two or even more than two simultaneous or nearly simultaneous selectable regulatory changes whose conjunction facilitates the building up of some new developmental, physiological, or morphological association of features. Species with relatively low population sizes might thus respond creatively to environmental challenges thanks to relatively "easy-to-come-by" combined regulatory changes of limited penetrance. These changes in gene action may be controlled by a much larger mutational target region than usually envisaged, a target region presented by non-protein-coding sequences in the neighborhood of a gene. Within this non-protein-coding target, many different mutations would have similar effects on the transcriptional expression of a given coding sequence. In this way, larger forms such as mammals may be able to make up for their limited population sizes and compete with the enormous bacterial populations in their ability to evolve sophisticated strategies at all levels of biological integration. The phenomenon is likely to have a developmental dimension, because a particular structure formed within non-protein-coding DNA sequences and their teleregulatory effects may both vary with the cell type and more generally the molecular background of the cell. SINEs and other retrotransposons might play a major role in combinatorial epigenetics. The proposed genetic/epigenetic partnership would in particular result from the fact that epigenetic changes in junk DNA that lead to a given regulatory effect would be expected to have a much lower requirement for sequence specificity and therefore for mutational specificity than classical mutations do. Adaptive epimutations are thus expected to occur much more often than adaptive mutations. By overcoming conditions created by small population sizes, the pathway outlined might have represented a breakthrough toward the evolution of the highest organisms.

Identical twins acquire epigenetic differences over their lifetime (Fraga et al., 2005). This observation attracts attention to an important potential implication of the inferred genetic/epigenetic partnership, namely, that it likely has a Lamarckian dimension. The case of the identical twins illustrates the finding that environmental factors can lead to epigenetic effects (Vickaryous and Whitelaw, 2005). Since at least some epigenetic modifications are cellularly inheritable, and since cellularly inheritable modifications can occur in the germ line, a pathway may be traced out here that leads to the inheritance of

acquired characteristics (Jablonka and Lamb, 1995; Roemer et al., 1997; Pal and Miklos, 1999). Most of these characteristics may be functionally nearly neutral or pathological. A minority of environmentally induced epigenetic changes would be expected to consist in adaptive physiological or morphological innovations, however, which consequently would be genetically assimilated. Along the particular pathway outlined, Lamarckian processes would thus be rare, and it could hardly be foreseen what form the adaptation would take. Such epigenetically acquired traits could nevertheless have great evolutionary impact. Philosophically, this version of Lamarckism would not be more finalistic than Darwinism.

Acknowledgements

E.Z. was supported through the Institute of Molecular Medical Sciences by the Sylvia Robb Charitable Trust. G.C. was supported by grants from the Centre National de la Recherche Scientifique, the Human Frontier Science Program Organization, the European Union FP 6 (Network of Excellence "The Epigenome" and STREP "3D Genome"), the Ministère de la Recherche (ACI BCMS 2004), the Agence Nationale de la Recherche, and by the Association pour la Recherche sur le Cancer.

References

Abranches, R., et al., 2000. Widely separated multiple transgene integration sites in wheat chromosomes are brought together at interphase. Plant J. 24, 713–723.

Adami, C., 2006. Evolution. Reducible complexity. Science 312, 61–63.

Allen, T.A., von Kaenel, S., Goodrich, J.A., Kugel, J.F., 2004. The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. Nat. Struct. Mol. Biol. 11, 816–821.

Antequera, F., Macleod, D., Bird, A.P., 1989. Specific protection of methylated CpGs in mammalian nuclei. Cell 58, 509–517.

Arnaud, P., Goubely, C., Pelissier, T., Deragon, J.M., 2000. SINE retroposons can be used in vivo as nucleation centers for de novo methylation. Mol. Cell. Biol. 20, 3434–3441.

Bacherm, C.P., et al., 2006. Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. Nat. Cell Biol. 8, 293–299.

Bantignies, F., Grimaud, C., Lavrov, S., Gabut, M., Cavalli, G., 2003. Inheritance of Polycomb-dependent chromosomal interactions in *Drosophila*. Genes Dev. 17, 2406–2420.

Bernardi, G., 2001. Misunderstandings about isochores. Gene 276, 3–13.

Bernardi, G., 2004. Structural and Evolutionary Genomics. Elsevier, Amsterdam.

Branco, M.R., Pombo, A., 2006. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol. 4, e138.

Bray, N.J., Buckland, P.R., Owen, M.J., O'Donovan, M.C., 2003. *Cis*-acting variation in the expression of a high proportion of genes in human brain. Hum. Genet. 113, 149–153.

Brem, R.B., Yvert, G., Clinton, R., Kruglyak, L., 2002. Genetic dissection of transcriptional regulation in budding yeast. Science 296, 752–755.

Bridgham, J.T., Carroll, S.M., Thornton, J.W., 2006. Evolution of hormone–receptor complexity by molecular exploitation. Science 312, 97–101.

Brown, K.E., Guest, S.S., Smale, S.T., Hahm, K., Merkenschlager, M., Fisher, A.G., 1997. Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. Cell 91, 845–854.

Brown, K.E., Baxter, J., Graf, D., Merkenschlager, M., Fisher, A.G., 1999. Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. Mol. Cell 3, 207–217.

Brown, J.M., et al., 2006. Coregulated human globin genes are frequently in spatial proximity when active. J. Cell Biol. 172, 177–187.

Burgess-Beusse, B., et al., 2002. The insulation of genes from external enhancers and silencing chromatin. Proc. Natl. Acad. Sci. U. S. A. 99 (Suppl. 4), 16433–16437.

Cao, R., et al., 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 298, 1039–1043.

Castillo-Davis, C., Mekhedov, S.L., Hartl, D.L., Koonin, E.V., Kondrashov, F.A., 2002. Selection for short introns in highly expressed genes. Nat. Genet. 31, 415–418.

Cavalli, G., Paro, R., 1998. Chromo-domain proteins: linking chromatin structure to epigenetic regulation. Curr. Opin. Cell Biol. 10, 354–360.

Cawley, S., et al., 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116, 499–509.

Chambeyron, S., Bickmore, W.A., 2004. Does looping and clustering in the nucleus regulate gene expression? Curr. Opin. Cell Biol. 16, 256–262.

Cheung, V.G., et al., 2003. Natural variation in human gene expression assessed in lymphoblastoid cells. Nat. Genet. 33, 422–425.

Coghlan, A., Eichler, E.E., Oliver, S.G., Paterson, A.H., Stein, L., 2005. Chromosome evolution in eukaryotes: a multi-kingdom perspective. Trends Genet. 21, 673–682.

Cowles, C.R., Hirschhorn, J.N., Altshuler, D., Lander, E.S., 2002. Detection of regulatory variation in mouse genes. Nat. Genet. 32, 432–437.

Csink, A.K., Henikoff, S., 1996. Genetic modification of heterochromatic association and nuclear organisation in *Drosophila*. Nature 381, 529–531.

Csink, A.K., Henikoff, S., 1998. Large-scale chromosomal movements during interphase progression in *Drosophila*. J. Cell Biol. 143, 13–22.

Cuénot, L., 1941. Invention et Finalité en Biologie. Flammarion, Paris.

Darwin, C., 1872. The Origin of Species, 6 ed. John Murray, London.

Dernburg, A.F., et al., 1996. Perturbation of nuclear architecture by long-distance chromosome interactions. Cell 85, 745–759.

D'Onofrio, G., 2002. Expression patterns and gene distribution in the human genome. Gene 300, 155–160.

Dorer, D.R., Henikoff, S., 1994. Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. Cell 77, 993–1002.

Dorer, D.R., Henikoff, S., 1997. Transgene repeat arrays interact with distant heterochromatin and cause silencing in *cis* and *trans*. Genetics 147, 1181–1190.

Duncan, I.W., 2002. Transvection effects in *Drosophila*. Annu. Rev. Genet. 36, 521–556.

Fraga, M.F., et al., 2005. Epigenetic differences arise during the lifetime of monozygotic twins. Proc. Natl. Acad. Sci. U. S. A. 102, 10604–10609.

Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., Cavalli, G., 2006. RNAi components are required for nuclear clustering of Polycomb group response elements. Cell 124, 957–971.

Harmon, B., Sedat, J., 2005. Cell-by-cell dissection of gene expression and chromosomal interactions reveals consequences of nuclear reorganization. PLoS Biol. 3, e67.

Henikoff, S., Furuyama, T., Ahmad, K., 2004a. Histone variants, nucleosome assembly and epigenetic inheritance. Trends Genet. 20, 320–326.

Henikoff, S., McKittrick, E., Ahmad, K., 2004b. Epigenetics, histone H3 variants, and the inheritance of chromatin states. Cold Spring Harbor Symp . Quant. Biol. 69, 235–243.

Hsieh, J., Fire, A., 2000. Recognition and silencing of repeated DNA. Annu. Rev. Genet. 34, 187–204.

Jablonka, E., Lamb, M.J., 1995. Epigenetic Inheritance and Evolution. The Lamarckian Dimension. Oxford University Press, Oxford.

Jaenisch, R., Bird, A., 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. 33, 245–254 (Suppl).

Jenuwein, T., Allis, C.D., 2001. Translating the histone code. Science 293, 1074–1080.

Jeong, S., Pfeifer, K., 2004. Shifting insulator boundaries. Nat. Genet. 36, 1036–1037.

Jones, P.A., Takai, D., 2001. The role of DNA methylation in mammalian epigenetics. Science 293, 1068–1070.

Kass, S.U., Goddard, J.P., Adams, R.L., 1993. Inactive chromatin spreads from a focus of methylation. Mol. Cell. Biol. 13, 7372–7379.

Kim, A., Dean, A., 2003. A human globin enhancer causes both discrete and widespread alterations in chromatin structure. Mol. Cell. Biol. 23, 8099–8109.

Kimura, H., Tada, M., Nakatsuji, N., Tada, T., 2004. Histone code modifications on pluripotential nuclei of reprogrammed somatic cells. Mol. Cell. Biol. 24, 5710–5720.

King, M.C., Wilson, A.C., 1975. Evolution at two levels in humans and chimpanzees. Science 188, 107–116.

Knight, J.C., 2004. Allele-specific gene expression uncovered. Trends Genet. 20, 113–116.

Kouskouti, A., Talianidis, I., 2005. Histone modifications defining active genes persist after transcriptional and mitotic inactivation. EMBO J. 24, 347–357.

Lenski, R.E., Ofria, C., Pennock, R.T., Adami, C., 2003. The evolutionary origin of complex features. Nature 423, 139–144.

Levine, A., Cantoni, G.L., Razin, A., 1991. Inhibition of promoter activity by methylation: possible involvement of protein mediators. Proc. Natl. Acad. Sci. U. S. A. 88, 6515–6518.

Levitsky, V.G., Podkolodnaya, O.A., Kolchanov, N.A., Podkolodny, N.L., 2001. Nucleosome formation potential of exons, introns, and Alu repeats. Bioinform. 17, 1062–1064.

Li, X., Noll, M., 1994. Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different *cis*-regulatory regions. Nature 367, 83–87.

Lindsay, H., Adams, R.L.P., 1996. Spreading of methylation along DNA. Biochem. J. 320, 473–478.

Ling, J., Ainol, L., Zhang, L., Yu, X., Pi, W., Tuan, D., 2004. HS2 enhancer function is blocked by a transcriptional terminator inserted between the enhancer and the promoter. J. Biol. Chem. 279, 51704–51713.

Ling, J., Baibakov, B., Pi, W., Emerson, B.M., Tuan, D., 2005. The HS2 enhancer of the beta-globin locus control region initiates synthesis of non-coding, polyadenylated RNAs independent of a *cis*-linked globin promoter. J. Mol. Biol. 350, 883–896.

Ling, J.Q., et al., 2006. CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. Science 312, 269–272.

Lo, H.S., et al., 2003. Allelic variation in gene expression is common in the human genome. Genome Res. 13, 1855–1862.

Locke, J., Kotarsky, M.A., Tartof, K.D., 1988. Dosage-dependent modifiers of position effect variegation in *Drosophila* and a mass action model that explains their effect. Genetics 120, 181–198.

Ma, Y., et al., 2005. DNA CpG hypomethylation induces heterochromatin reorganization involving the histone variant macroH2A. J. Cell. Sci. 118, 1607–1616.

Makalowki, W., Mitchell, G.A., Labuda, D., 1994. Alu sequences in the coding regions of mRNA: a source of protein variability. Trends Genet. 10, 188–193.

Morris, K.V., Chan, S.W.L., Jacobsen, S.E., Looney, D.J., 2004. Small interfering RNA-induced transcriptional gene silencing in human cells. Science 305, 1289–1292.

Mutskov, V., Felsenfeld, G., 2004. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. EMBO J. 23, 138–149.

Mutskov, V.J., Farrell, C.M., Wade, P.A., Wolffe, A.P., Felsenfeld, G., 2002. The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. Genes Dev. 16, 1540–1554.

Nekrutenko, A., Li, W.H., 2001. Transposable elements are found in a large number of human protein-coding genes. Trends Genet. 17, 619–621.

Ng, N.H., Robert, F., Young, R.A., Struhl, K., 2003. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol. Cell 1, 709–719.

Nikiforova, M.N., Stringer, J.R., Blough, R., Medvedovic, M., Fagin, J.A., Nikiforov, Y.E., 2000. Proximity of chromosomal loci that participate in radiation-induced rearrangements in human cells. Science 290, 138–141.

Noma, K., Cam, H.P., Maraia, R.J., Grewal, S.I., 2006. A role for TFIIIC transcription factor complex in genome organization. Cell 125, 859–872.

Osborne, C.S., et al., 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. Nat. Genet. 36, 1065–1071.

Pal, C., Miklos, I., 1999. Epigenetic inheritance, genetic assimilation and speciation. J. Theor. Biol. 200, 19–37.

Pastinen, T., et al., 2003. A survey of genetic and epigenetic variation affecting human gene expression. Physiol. Genomics 16, 184–193.

Pearson, H., 2006. What is a gene? Nature 441, 399–401.

Polak, P., 2006, "Master thesis", personal communication.

Polak, P., Domany, E., 2006. Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. BMC Genomics 7, 133.

Poux, S., Horrard, B., Sigrist, C.J., Pirrotta, V., 2002. The *Drosophila* trithorax protein is a coactivator required to prevent re-establishment of polycomb silencing. Development 129, 2483–2493.

Rangwala, S.H., Richards, E.J., 2004. The value-added genome: building and maintaining genomic cytosine methylation landscapes. Curr. Opin. Genet. Dev. 14, 686–691.

Raser, J.M., O'Shea, E.K., 2005. Noise in gene expression: origins, consequences, and control. Science 309, 2010–2013.

Richards, E.J., Elgin, S.C., 2002. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell 108, 489–500.

Roemer, I., Reik, W., Dean, W., Klose, J., 1997. Epigenetic inheritance in the mouse. Curr. Biol. 7, 277–280.

Roix, J.J., McQueen, P.G., Munson, P.J., Parada, L.A., Misteli, T., 2003. Spatial proximity of translocation-prone gene loci in human lymphomas. Nat. Genet. 34, 287–291.

Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S., Elowitz, M.B., 2005. Gene regulation at the single-cell level. Science 307, 1962–1965.

Rubin, H., 1990. The significance of biological heterogeneity. Cancer Metastasis Rev. 9, 1–20.

Sabl, J.F., Henikoff, S., 1996. Copy number and orientation determine the susceptibility of a gene to silencing by nearby heterochromatin in *Drosophila*. Genetics 142, 447–458.

Schmitt, S., Prestel, M., Paro, R., 2005. Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev. 19, 697–708.

Schreiber, S.L., Bernstein, B.E., 2002. Signaling network model of chromatin. Cell 111, 771–778.

Schubeler, D., et al., 2004. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev. 18, 1263–1271.

Silva, A.J., White, R., 1988. Inheritance of allelic blueprints for methylation patterns. Cell 54, 145–152.

Snowden, A.W., Gregory, P.D., Case, C.C., Pabo, C.O., 2002. Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. Curr. Biol. 12, 2159–2166.

Sollars, V., Lu, X., Xiao, L., Wang, X., Garfinkel, M.D., Ruden, D.M., 2003. Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. Nat. Genet. 33, 70–74.

Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., Flavell, R.A., 2005. Interchromosomal associations between alternatively expressed loci. Nature 435, 637–645.

Su, R.C., Brown, K.E., Saaber, S., Fisher, A.G., Merkenschlager, M., Smale, S.T., 2004. Dynamic assembly of silent chromatin during thymocyte maturation. Nat. Genet. 36, 502–506.

Su, R.C., Sridharan, R., Smale, S.T., 2005. Assembly of silent chromatin during thymocyte development. Semin. Immunol. 17, 129–140.

Sutherland, H.G., et al., 2000. Reactivation of heritably silenced gene expression in mice. Mamm. Genome 11, 347–355.

Tolhuis, B., et al., 2006. Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster*. Nat. Genet. 38, 694–699.

Torok, T., Gorjanacz, M., Bryant, P.J., Kiss, I., 2000. Prod is a novel DNA-binding protein that binds to the 1.686 g/cm(3) 10 bp satellite repeat of *Drosophila melanogaster*. Nucleic Acids Res. 28, 3551–3557.

Trifonov, E.N., 2004. Tuning function of tandemly repeating sequences: a molecular device for fast adaptation. In: Wasser, S.P. (Ed.), Evolutionary Theory and Processes: Modern Horizons, Papers in Honour of Eviatar Nevo. Kluwer Academic Publishers, The Netherlands, pp. 115–138.

Turker, M.S., 1999. The establishment and maintenance of DNA methylation patterns in mouse somatic cells. Semin. Cancer Biol. 9, 329–337.

Vickaryous, N., Whitelaw, E., 2005. The role of early embryonic environment on epigenotype and phenotype. Reprod. Fertil. Dev. 17, 335–340.

Vinogradov, A.E., 2005. Dualism of gene GC content and CpG pattern in regard to expression in the human genome: magnitude versus breadth. Trends Genet. 21, 639–643.

Waddington, C.H., 1942. The epigenotype. Endeavor 1, 18–20.

West, A.G., Huang, S., Gaszner, M., Litt, M.D., Felsenfeld, G., 2004. Recruitment of histone modifications by USF proteins at a vertebrate barrier element. Mol. Cell 16, 500–502.

Xu, N., Tsai, C.L., Lee, J.T., 2006. Transient homologous chromosome pairing marks the onset of X inactivation. Science 311, 1149–1152.

Yates, P.A., Burman, R.W., Mummaneni, P., Krussel, S., Turker, M.S., 1999. Tandem B1 elements located in a mouse methylation center provide a target for de novo DNA methylation. J. Biol. Chem. 274, 36357–36361.

Yates, P.A., Burman, R., Simpson, J., Ponomoreva, O.N., Thayer, M.J., Turker, M.S., 2003. Silencing of mouse Aprt is a gradual process in differentiated cells. Mol. Cell. Biol. 23, 4461–4470.

Yusufzai, T.M., Tagami, H., Nakatani, Y., Felsenfeld, G., 2004. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol. Cell 13, 291–298.

Zhao, H., Dean, A., 2004. An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. Nucleic Acids Res. 32, 4903–4919.

Zilberman, D., Henikoff, S., 2005. Epigenetic inheritance in *Arabidopsis*: selective silence. Curr. Opin. Genet. Dev. 15, 557–562.

Zuckerkandl, E., 1963. Perspectives in molecular anthropology. In: Washburn, S.L. (Ed.), Classification and Human Evolution. Aldine Publishing Co., Chicago.

Zuckerkandl, E., 1964. Controller-gene diseases: the operon model as applied to beta-thalassemia, familial fetal hemoglobinemia and the normal switch from the production of fetal hemoglobin to that of adult hemoglobin. J. Mol. Biol. 37, 128–147.

Zuckerkandl, E., 1968. Hemoglobins, Haeckel's "biogenetic law", and molecular aspects of development. In: Rich, Alexander, Davidson, Norman (Eds.), Structural Chemistry and Molecular Biology. W.H. Freeman and Co., San Francisco, pp. 256–274.

Zuckerkandl, E., 1974. A possible role of "inert" heterochromatin in cell differentiation. Action of and competition for "locking molecules". Biochimie 56, 937–954.

Zuckerkandl, E., 1981. A general function of noncoding polynucleotide sequences. Mass binding of transconformational proteins. Mol. Biol. Rep. 7, 149–158.

Zuckerkandl, E., 1983. Topological and quantitative relationships in evolving genomes. In: Hélène, C. (Ed.), Structure, Dynamics, Interactions and Evolution of Biological Macromolecules. D. Reidel Publishing Co, pp. 395–412.

Zuckerkandl, E., 1997. Junk DNA and sectorial gene repression. Gene 205, 323–343.

Zuckerkandl, E., 1999. Sectorial gene repression in the control of development. Gene 238, 263–276.

Zuckerkandl, E., 2002. Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine. Genetica 115, 105–129.

Zuckerkandl, E., Pauling, L., 1965. Evolutionary divergence and convergence in proteins. In: Bryson, V., Vogel, H.J. (Eds.), Evolving Genes and Proteins. Academic Press, New York, pp. 97–165.