Is ectopic expression caused by deregulatory mutations or due to gene-regulation leaks with evolutionary potential?

Francisco Rodríguez-Trelles,[1,2]* Rosa Tarrío,[1,2] and Francisco J. Ayala[1]

Summary

It has long been thought that gene expression is tightly regulated in multicellular eukaryotes, so that expression profiles match functional profiles. This conception emerged from the assumption that gene activity is synonymous with gene function. This paradigm was first challenged by comparative protein electrophoresis studies showing extensive differences in expression patterns among related species. The paradigm is now being challenged by evolutionary transcriptomics using microarray technologies. Most gene expression profiles display features that lack any obvious functional significance. The so-called ''ectopic'' expression refers to the expression of genes at times and locations where the target gene is not known to have a function. However, ectopic expression might be associated with genuine function even if this function is not essential or has yet to be ascertained. Alternatively, ectopic expression might come about as a superfluous by-product of regulatory systems, which would call for a revision of prevailing ideas about the specificity of gene regulation. We herein review available evidence for ectopic expression and the hypotheses proposed for its origin and evolution. We propose that ectopic expression must be regarded as part of an integrated phenotypic whole. It seems likely that ectopic expression represents a leak in the evolution of regulatory systems, but one that is endowed with considerable evolutionary possibilities. ***BioEssays*** **27: 592–601, 2005. © 2005 Wiley Periodicals, Inc.**

[1]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA.
[2]Grupo de Xenética Evolutiva, Departamento de Anatomía Patolóxica e Ciencias Forenses, Universidade de Santiago de Compostela, Santiago, Spain.
Funding agency: F.R-T and R.T. have received support from contracts Ramón y Cajal and F.I.S. G03/184, from the Ministerio de Ciencia y Tecnología and Ministerio de Sanidad (Spain), respectively. Research supported by NIH Grant GM42397 to F.J.A.
*Correspondence to: Francisco Rodríguez-Trelles, c/o Francisco J. Ayala, Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525. E-mail: ftrelles@usc.es
DOI 10.1002/bies.20241
Published online in Wiley InterScience (www.interscience.wiley.com).

Introduction

Current understanding of the genotype–phenotype relations in multicellular eukaryotes rests upon two assumptions, namely that: (1) gene activity is identical to gene function, so that transcriptional profiles are faithful representations of functional scopes, and (2) gene expression specificity is a highly regulated, essentially error-free process.[1] These two assumptions are often advanced in a circular manner as if they would necessitate each other. The functional demands of gene expression require that expression be accurately controlled and, conversely, gene expression must always meet a function, given the specificity with which it is effected.

The assumption 'gene activity equals gene function' emerged as a logical corollary of the one-to-one correspondences 'gene–trait' and 'trait–function' asserted, respectively, by Mendelian genetics and adaptive evolution. The premise was bolstered by the articulation of the mutation–selection balance model of classical population genetics,[2,3] by which nonfunctional gene products are weeded out by natural selection. The quest for understanding the stable biochemical and structural identities that arise at the higher levels of physiological and developmental organization called for a model of specificity of gene action.[4] Following the formulation of the 'one gene–one protein' hypothesis, the discovery of the double helix provided an informational model of specificity embodied in the uniqueness of a nucleotide sequence. Specificity was perceived, at this point, as a static feature materialized in a linear molecule. The 'central dogma', i.e. the notion that genetic information flows from DNA to RNA to protein, freed informational specificity from its DNA corset. The high-fidelity processes of transcription and translation ensured that the static informational specificity of DNA could be accurately converted into the dynamic specificity typified in the stereochemical configuration of protein. Coupled with the diffusion of the gene products, these processes allowed informational specificity to be amplified and delivered far from the DNA template. However, which protein and in what amount it would be present in a certain tissue and at a certain time could not be a matter of chance; regulatory specificity was expected.

The conceptual leap from an informational model of gene specificity to a functional one gained general acceptance after the formulation of the operon model of gene regulation.[5] Many observations of *Drosophila* experimental genetics fit this paradigm well. Genes are endowed with binary switches that can be either 'on' or 'off'. Development is defined as the outcome of programmed differential gene activating and silencing. Albeit the same genetic information is present within every specialized cell, genes not involved in a cell's distinctive physiological and/or developmental commitment are turned off, with the same consequences as if they had been excised out.[6–8] Homeotic mutants are par excellence examples illustrating the dramatic consequences of turning on genes unspecifically.[9] The spatiotemporal activity patterns of genes are atomized into unitary traits, so that every feature of a gene's transcriptional profile is explained as the result of functional optimization. Thus, experimentally induced ectopic expression methods were originally conceived as a disrupting approach; i.e. a mode of investigating a gene's function by analyzing the phenotypic consequences of driving its expression in tissues in which it is normally not active.

The emerging picture of informational specificity translated into functional specificity, which we have just described, had some, now obvious, limitations. Bacterial molecular genetics did not contend with the intricacies of high gene numbers interspersed with vast stretches of non-coding DNA, which has recently been found to be extensively transcribed, compact packing of large genomes and cellular specialization, all features that characterize multicellular eukaryotes. *Drosophila* experiments focused on a few gene families that regulate major aspects of body pattern,[10,11] the so-called genetic 'toolkit' for animal development.[7] These are developmentally essential genes, not representative of a typical transcribed sequence.[9] Induced ectopic expression studies and site-directed mutagenesis assays are generally aimed at producing data consistent with the extant regulatory paradigm, and often fail to take into account negative results, i.e. outcomes in which misexpression does not produce a recognizable phenotype.

In this essay, we review evidence that challenges the two long-standing assumptions stated above (identity of gene activity and gene function, and precise regulation of gene expression) on the basis of recent experimental evidence from novel evolutionary transcriptomics approaches,[12–14] and theoretical ideas inspired by the emerging picture about the rationale and the mechanics of the regulation of eukaryotic genes' expression.[15,16]

Models of the extent of the selective regulation of gene activity constrain our understanding of what genes do, and the converse. We contend that contemplating gene expression as a highly regulated process, functionally fine-tuned down to the last detail, has largely led us to overlook situations in which gene activity does not convey any obvious functional significance. For instance, combined search for 'ectopic expression' and 'evolution' in popular literature databases yields zero citations. Often dubbed 'illegitimate', 'spurious', or 'superfluous' expression, because it takes place in cell types different from those where the subject gene is known to be normally active—so presumptively effecting a function—ectopic expression is emerging as a widespread phenomenon in the highly specialized cells of multicellular eukaryotes, which has important evolutionary significance.

Gene activity may not generally be synonymous with gene function

Early hints that gene activity might not always be synonymous with gene function in complex multicellular eukaryotes can be traced back to the interspecific comparative electrophoretic assays of known protein products actively pursued since the late nineteen seventies.[17–21] Dickinson's[17] prospect for changes in regulatory patterns using six enzymes across 14 larval and adult tissues of $\sim$30 Hawaiian picture-wing *Drosophila* species revealed extensive interspecific differences in expression patterns (conservatively affecting $>$30% of the individual traits included in the study), even among closely related species.[17] The highest expression levels observed were evolutionarily conserved; i.e., they occurred in the same tissue(s) in all the species where obvious functional significance could be construed. For example, alcohol dehydrogenase was always present in the fat body, and aldehyde oxidase-2 in the mid-gut. Most interspecific differences were concentrated in 'secondary sites of expression', i.e. tissues where none of the species exhibited high activity of the subject enzyme, suggesting that much of the interspecific variability in expression patterns was not adaptive.

Dickinson's[17] assays were not designed to test the functional significance of this dramatic interspecific variation. Moreover, there was no detailed physiological information about the enzymes' roles in the species-specific locations. Consequently, Dickinson was unable to advance any conclusive interpretation of the observed patterns, except for the general observation that the expression profiles seemed to diverge strikingly fast.[15,17–22] Unexpectedly high levels of variation in enzyme expression patterns were subsequently reported in several other *Drosophila* studies.[18–21] In addition, studies reporting analogous observations to Dickinson's[17] in vertebrate neurotransmitters also failed to establish any physiologically relevant role in secondary sites of expression from extensive data on neuropeptidergic function.[23]

Further signs that a gene's activity might not always imply function emerged from the observation of homozygotes for null alleles, namely, alleles that do not produce any distinguishable gene product, but survive perfectly well,[24,25] and from the absence of detectable phenotypic effect often encountered in many genetic knockouts.[26–29] These two cases question the functional significance of the genes' usual

products—although the generality of this conclusion is constrained by the observation that transcription factors, membrane receptors and other macromolecular complexes often exhibit haplo-insufficient phenotypes.[30,31] A similar challenge arises in the case of expressed pseudogenes, i.e. gene copies that (generally) do not produce any functional gene product.[32] For instance, human myosin XVBP and L-threonine 3-dehydrogenase genes, located respectively at 17q25 and 8p23-22, are unprocessed pseudogenes that actively produce frame-shifted and/or truncated products in various tissues.[33,34]

Genome-wide indications that nonfunctional expression might be a widespread natural phenomenon in the specialized cells of multicellular eukaryotes has emerged recently, with the advent of powerful microarray technologies. Microarray experiments enable studies of an organism's transcriptome, i.e., the whole set of transcripts and their relative levels of expression in a cell or tissue type (Fig. 1). Using these methods, Khaitovich et al.[13] have failed to reject the prediction derived from a strictly neutrality model for polygenic quantitative traits that, if evolutionary changes in the levels of gene expression are caused by selectively equivalent alleles, the variance of expression levels among populations or species should be proportional to the time since their common ancestry.[35,36] Khaitovich et al.[13] have assayed the transcript expression level differences of 12,000 genes in samples of prefrontal cortex from humans and three other primates, chimp, orang utang and macaque. They plot the average squared difference between log-expression levels of all genes with detectable expression against molecular estimates of the species' age, and obtain an 'approximate' linear rate of accumulation of expression differences over evolutionary time (see Fig. 2). Analogous linear relationships are obtained for the liver, after extending the analyses to three mice species, and when comparing different tissues within the same individual—whether a human, chimp or mouse. But, as noted by Khaitovich et al.,[13] a clock-like behaviour of the among-lineages variance of expression could also result from selection; e.g., if the optimum phenotype wanders erratically in a Brownian motion-like fashion.[37–39] Khaitovich et al.[13]

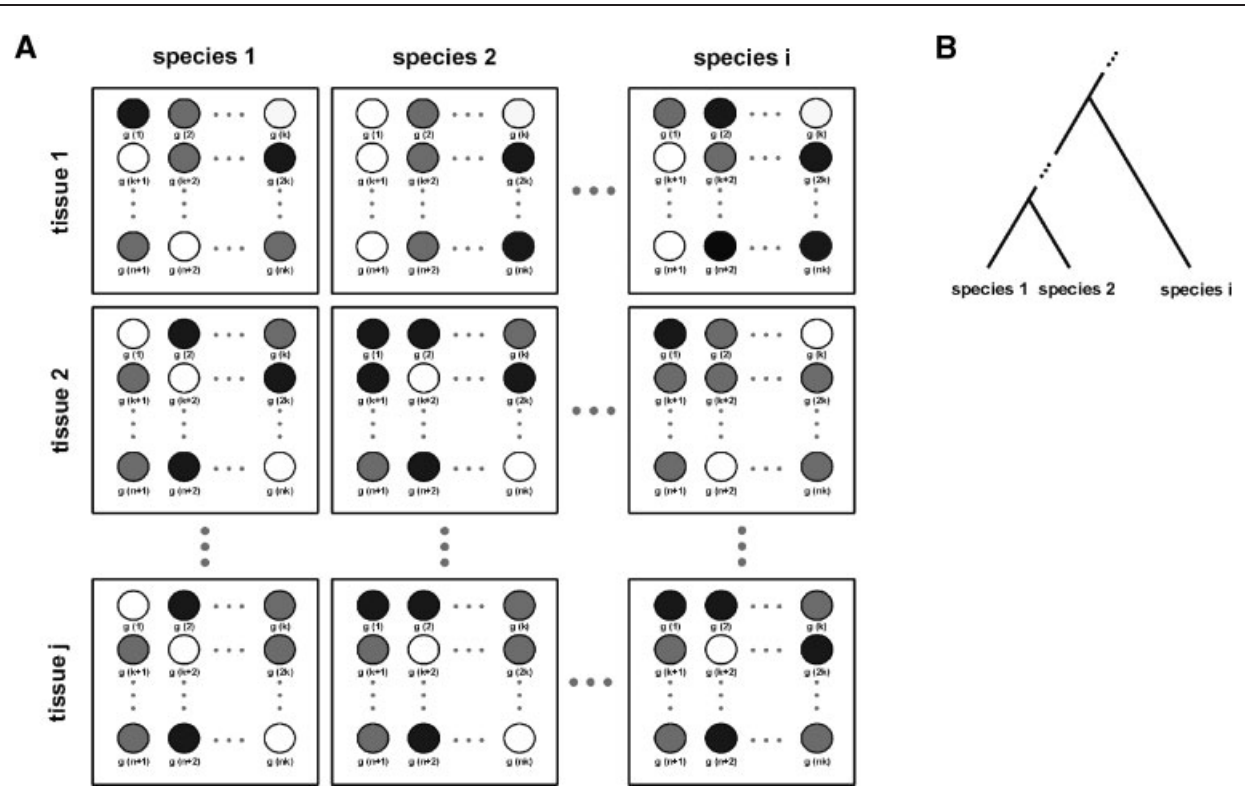

Figure 1. A: Representation of an evolutionary microarray experiment comparing expression profiles for individual genes (1 to *nk*) across *j* different tissues, under defined conditions, across *i* different species. The species are known to be related phylogenetically as in **B**. For simplicity, only three quantitative expression levels are represented: high (black), intermediate (gray) and non-detectable (white). For example, the expression level of gene 1 in tissue 1 is not evolutionary conserved (it is high in species 1, undetectable in species 2 and intermediate in species 3), while the expression level of gene 2 is conserved in tissue j (it is high in all species). Neutrality models predict that the amount of divergence in a gene's transcriptional profile among different species is consistent with the species' phylogeny.

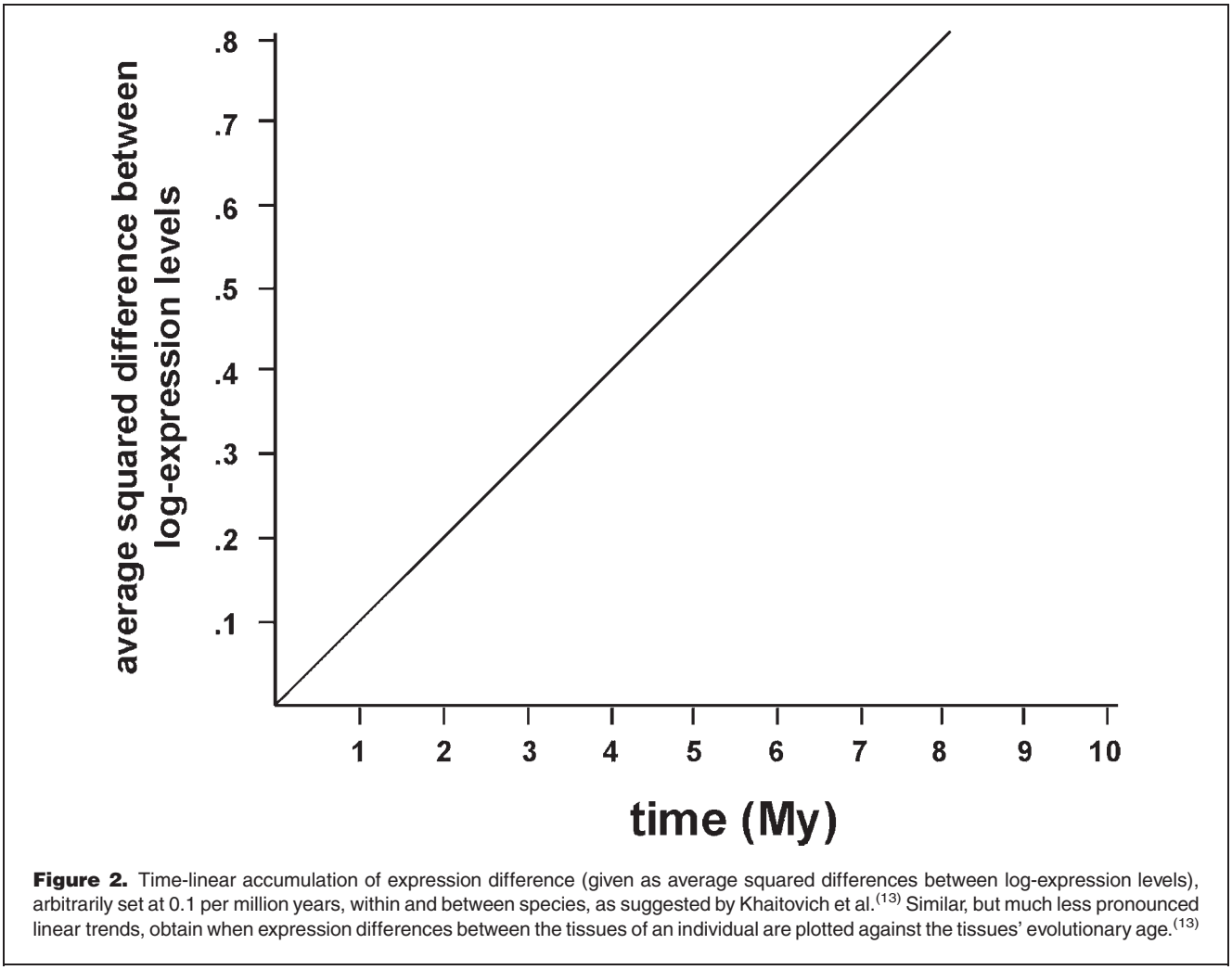

Figure 2. Time-linear accumulation of expression difference (given as average squared differences between log-expression levels), arbitrarily set at 0.1 per million years, within and between species, as suggested by Khaitovich et al.[13] Similar, but much less pronounced linear trends, obtain when expression differences between the tissues of an individual are plotted against the tissues' evolutionary age.[13]

therefore conducted an additional set of experiments. They found that the variation in gene expression between humans and chimps is positively correlated with the variation in gene expression within humans, which agrees with the results of a previous survey using species of teleost fish.[40] In addition, the rates of expression divergence between humans and chimps do not significantly differ between intact genes and expressed pseudogenes. These and other results showing differences in gene expression profiles that follow phylogenetic relationships among closely related species,[17,22,41,42] lead Khaitovich et al.[13] to conclude that the majority of expression differences observed within and between species are of little or no adaptive significance, being ultimately determined by chance. They predicate this outcome on the fact that the transcriptome level is closer to the genotype level, where the variation is generated, than to the phenotype level, where the variation is filtered by natural selection. The specific reasons why proximity to the genotype level should imply that transcript levels are only 'secondary' phenotypic attributes (while morphology and behavior would be primary features) are left unstated.

The rationale of Khaitovich et al.[13] for inferring absence of functional significance needs scrutiny. First, the organisms compared are not grown in a defined environment, which questions the validity of Khaitovich et al's[13] data as strict tests of any genetic hypothesis. One could reasonably argue that the observed expression-level changes are simply due to environmental differences.[37] Besides this, Khaitovich et al[13] base their conclusion of a time-linear behaviour of the among-lineages variance of expression on eye-inspection of their plots. Yet the perceived correlation between expression divergence and time could be inflated owing to the few pair-wise comparisons used in their study (involving as few as four taxa), which, moreover, are not independent since they are based on shared portions of the phylogenetic tree.[43] Linearity is further emphasized by confidence intervals which only take into account the variance of the estimates of expression levels—thus neglecting the uncertainty introduced by the stochastic

variation of transcriptome evolution itself—and ignore the scattering of primate divergence dates associated with molecular studies different from the one they use as reference.[44–46] Khaitovich et al's test would seem to lack sufficient statistical power to justify by itself acceptance of the mutation-drift model. In any case, consistency with drift would only mean that drift can not be rejected, which is not the same as asserting that selection has not played a central role. Their test of positive selection based on expressed pseudogenes as nonfunctional features is unwarranted, for these sequences are increasingly known to carry functions.[32,47]

Other studies suggest that the evolution of expression-level differences reflects locus, lineage and locus–lineage interaction effects. Comparative studies analysing genome-wide expression differences on a gene-wise basis find that genes can be classified into different evolutionary modes, e.g. stabilizing, directional and balancing selection in addition to drift, according to their ratios of intraspecific to interspecific expression variation patterns.[40,41,48,49] In contrast to Khaitovich et al.,[13] a microarray survey of gene activity variation among members of the *Drosophila melanogaster* subgroup found most transcription differences to be under stabilizing selection.[41] A greater effectiveness of selection in *Drosophila* species than in mammals (primates and rodents) might be expected, because the former have larger effective population sizes than the latter. Stable expression patterns in *Drosophila* have allowed the identification of genes encoding transcription factors, even though the downstream targets of these genes display expression variation;[41] expression divergence is positively correlated with the rate of protein evolution for 195 genes that are expressed in at least one species out of *D. melanogaster* and *D. simulans*.[49] In the teleost fish *Fundulus*, some gene expression patterns reflect the environments in which the species have evolved regardless of their phylogenetic relationships.[40] Aerobic energy metabolism genes and neuronal function-related genes display upregulated expression profiles in the anterior cingulated cortex of humans and chimps relative to gorillas and macaques,[42] and overexpression involving several gene functional classes and tissues distinguish humans from African great apes.[50–52] Besides these studies, there is an increasing number of examples evincing that natural selection operates on allelic variation in the *cis*-regulatory sequences that control transcription.[53,54] Lumping all genes together irrespective of their functions will tend to emphasize more abundant transcript genes, which may preferentially encompass certain functional classes, such as metabolic-function genes, certain ubiquitous signalling functions, or cytoskeletal proteins. Similarly, not taking into account genes' family structure may lead to overemphasize duplicate genes, which are known to increase gene expression diversity within and between species compared to single-copy genes.[55–57] Transcriptome comparisons of present-day species average over their entire period of divergence, which is likely to mask short episodes of increased and decreased rates driven by selection.[37] For example, Thorpe et al[19] found that the degree of enzyme pattern variation was about 60% higher in Hawaiian *Drosophila* than among species of the *D. virilis* group. They attributed this difference to bottlenecks and founder effects during the radiation of Drosophilids in Hawaii, which may have resulted in the chance fixation of many neutral variants. Recently, the ratio of brain and liver expression divergence to neutral sequence divergence has been found to be about twice as large in hominids (human and chimp) as in mice (*M. domesticus* and *M. spretus*). This difference has been attributed to a reduction in the effectiveness of natural selection in hominids owing to their lower effective population size.[54] This observation corroborates that the proportion of expression changes that are under natural selection varies among different lineages. Comparisons of genome-wide expression averages across contemporary organisms, such as that done by Khaitovich et al.,[13] is likely to yield a history of change that cannot be distinguished from that of a random processes.

Neutrality is a theory about the relative effects of different alleles on reproductive success. The observation that detectable transcript levels in a cell type or tissue have no discernible fitness effects, indicates that performance is not affected by the transcript levels under consideration. It does not follow that the encoding gene lacks a function in that tissue. For instance, the gene's activity could be only constrained by a minimum transcript level, such that amounts of transcript over that threshold would not have functional consequences. Therefore, strictly speaking, the results of Khaitovich et al.[13] would not disprove that gene activity equals gene function. This equivalence has been challenged by Yanai et al.[12] in a comparison of the expression profiles of 1,350 orthologous gene pairs from human and mouse across sixteen different tissues. They observe that expression of human genes changes, from one to another tissue, in a manner that cannot be anticipated from the corresponding expression shifts of their orthologous genes in the mouse, even though the function remains the same in humans and mice. Extensive expression differences between the two species are frequently of the type 'all-or-nothing' (see, for example, the expression of gene 1 in tissue 1 across species 1 and 2 in Fig. 1); for many combinations of gene and tissue, expression is detected in one member of the orthologous pair but not in the other. It seems unlikely that, in these cases, gene activity in the species and tissues in which transcription is observed in only one species has functional significance. Among examples highlighted by Yanai et al.,[12] the ENO2 gene encodes a neural enolase known to play a role in the nervous system. ENO2 high expression levels are conserved between human and mouse cerebellum, amygdala and the dorsal root ganglion. But ENO2 is also active in the human uterus, where most likely it does not perform any function,

given that ENO2 expression could not be detected in the mouse uterus.[12]

Yanai et al.[12] explain these observations as the widespread occurrence of mutations for ectopic expression, which become fixed in populations by random drift because they are inconsequential to fitness. A caveat to this explanation is that finding a given gene transcriptionally 'off' in a cell type might not necessarily indicate that that gene's function is not performed in that cell type. It is as yet unknown how compensatory gene functions vary. For instance, it might happen that a particular gene function that is not expressed in organism A compared to organism B, is compensated by paralogues, or even genes that bear little sequence relatedness but happen to have similar functional attributes, which are expressed at higher levels in A.[58]

Yanai et al.'s[12] interpretation fits in with known properties of binding sites for the many different transcription factors, which are typically short (6 to 10 base pairs long) and imprecise, such that new, presumably often spurious binding sites, with the potential for novel regulatory interactions, can be generated (or lost) by random mutation at high frequencies.[59,60] A survey of *cis*-regulatory variation in humans has uncovered extensive polymorphisms, even though the study was exclusively circumscribed to experimentally verified functional variants (i.e. with detectable phenotypic effects) of protein-coding transcript levels.[61] Several cases of polymorphic temporal and spatial regulation were found; for example, one at the FY locus that affects expression in erythroid cells but not in other tissues,[61,62] and others at the HGB2 locus that alters the time course of expression, resulting in variable fetal hemoglobin expression among adults.[61,63]

According to Yanai et al.'s[12] view, ectopic expression would be an accidental phenomenon caused by disabling, yet harmless, mutations irrelevant to an elaborate controlling apparatus. An alternative, yet not mutually exclusive explanation, is that ectopic expression could be inherent to the regulatory system: a consequence of internal constraints evolved as acompromise between workable regulatory schemes and viable DNA–protein spatial configurations within the eukaryotic nucleus. A great deal of progress in the understanding of the rationale and mechanics of gene control made in recent years[6–8] strongly suggests that the regulation of gene expression is inherently a leaky process.

Molecular basis of the specificity of gene regulation

The specificity of gene activity is accomplished by cooperative regulation at two levels: (i) restricting the ability of transcriptional activators to access their target DNA sequences, and (ii) modulating particular combinations of transcription factors. The first is achieved by modulating the structural configuration of the chromosomal environment in which genes are embedded[64–66] and the localization of genes within the nucleus.[1,67] DNA can be either tightly packed into chromatin, or decondensed via chromatin-remodeling complexes. To be transcribable, a gene must be both in decondensed, 'open' chromatin state and localized in the same nuclear compartment as their regulatory effectors.

The second level of cooperative regulation occurs because there are a limited number of transcription factors. Their regulatory potential is greatly expanded because of the transcription factors' ability to associate following combinatorial principles, so as to form multiprotein transcription complexes.[15,68–73] Transcription factor complexes target clusters of binding sites or enhancers that are dispersed across *cis*-regulatory regions, i.e. stretches of DNA that occur far removed—even intermingled with those of other genes—or within the genes that they regulate. A typical enhancer contains of the order of ten binding sites for at least three different transcription factors.[6] Enhancers operate as autonomous units or modules that act in a distance-independent manner. Individual modules direct or repress transcription in specific cell types and at particular times, each influencing just a discrete aspect of the overall transcriptional profile.[74–76] Different modules can direct composite patterns of gene expression when linked within a common *cis*-regulatory region.[77]

However, the following applies. (1) At any given physiological/developmental stage or in any cell type, genes exposed to the transcriptional apparatus in decondensed chromatin status, are not necessarily functionally related to one another.[16,78] (2) *Cis*-regulatory regions from genes with distinct transcriptional programs can exhibit enhancer modules for the same set of transcription factors and, the other way around, individual modules can exert effects over large distances, even between homologous chromosomes, thus being potentially capable of influencing transcriptional activities from remote locations.[6] The recurrent use of a limited number of elements, in varying combinatorial regulatory schemes, can theoretically generate a great deal of interconnectivity among otherwise unrelated transcriptional profiles,[15,79] such that it must be nearly impossible that any given aspect of an expression pattern can be optimized without imposing limits on others.

It seems extremely unlikely that each gene can have its own unique regulatory mechanism. Rather, it is expected that, in regulating their bona fide targets, many transcription factor complexes interact fortuitously and/or less specifically with the *cis*-regulatory modules of other functionally unrelated, yet co-exposed (in open-chromatin state) genes.[72] A recent transcriptional profiling using microarrays[16] has shown that 20% of the genes in the *Drosophila* genome fall into groups of 10–30 contiguous genes which display similar expression profiles although they do not exhibit any obvious functional relationship. Frequently, one or two genes in a group display high levels of differential expression, which suggests that the remainder transcriptional activity from the group is a

secondary effect, triggered because the entire group happened to be encompassed by the (minimum, given architectural constraints) region of decondensed chromatin that makes possible the expression of these genes.[16]

If the fluke activities resulting from such a leaky regulatory scheme would have none or little biological significance, it follows that one should observe ectopic expression everywhere, as suggested by evolutionary transcriptome studies.[12,13,17,23] The observation of widespread occurrence of ectopic expression is consistent with recent claims that insulator sequences, boundary elements that putatively prevent intergenic cross-talking problems by limiting the action of enhancers, seem to be less common than initially thought.[80] Accordingly, insulators would have evolved only in highly specialized contexts, i.e. loci displaying high density of coding or regulatory information, in which regulatory programs must likely be deployed very precisely.

Newly devised oligonucleotide microarray 'tiling' experiments that assay transcription at regular intervals throughout the genome—thus providing an unbiased estimate of transcription with respect to the location of known genes—have corroborated early hints that the transcriptome is much larger than first considered.[81–85] These methods reveal ten times as much transcription in human chromosomes 21 and 22 than is accounted for by known or predicted exons,[82,83] and widespread transcription has similarly been detected in other species.[85] The extra amounts of transcription thus uncovered come from outside the boundaries of known genes, including intergenic regions, introns of known genes and sequences antisense to known transcripts.[81,85] The meaning of this new, so dubbed 'dark-matter' transcription, is largely unknown, but there exists the possibility that much of it is the reflection of spurious binding sites randomly occurring in chromatin regions that happen to be open for transcription owing to constraints of the regulatory system. This view might be supported by the observation that many of the tiling predicted transcripts appear to be expressed at low levels,[82,84] much as in Dickinson's[17] secondary spots of activity. Ectopic gene expression might thus just be the tip of an iceberg of superfluous transcription affecting the whole genome.

Evolutionary implications of ectopic expression

In discussing ectopic gene activity, the notion of superfluous expression may be somewhat misleading. The contention is not that a gene product is completely devoid of function. Rather, the argument is that gene products may not necessarily have a function everywhere that they are expressed. In addition, it should be noticed that the term 'superfluous' is often meant solely to signify inability to ascertain a function. Establishing a function might require the testing of vast multidimensional arrays of conditions. Not being able to ascertain a function is, by no means, proof that no function exists.[15] Be that as it may, evolutionary transcriptomics suggests that gene expression profiles frequently display features that are of little significance to the fitness of the organism. An alternative to the hypothesis, that the rapid time-linear accumulation of transcriptome differences between species is caused by widespread and superfluous ectopic expression, is the 'marginal benefit' hypothesis.

Originally articulated to account for the observation that a majority of yeast genes appear to be dispensable—at least under laboratory conditions[86]—the marginal benefit hypothesis holds that gene activities are always functional, yet there are cell types and tissues where a gene product may exist only to fine-tune functions carried out primarily by other genes, or it may be involved in processes that are nonessential or not very important. The most-extreme form of this hypothesis would not allow ectopic expression. A cell's RNA turnover is mostly due to mRNA transcription. Transcription can consume a substantial portion of a cell's energy budget, up to 10% in mammal hepatocytes.[87] If widespread ectopic expression were gratuitous, it would represent a metabolic drain to the cell. This burden could be much greater if ectopic transcripts are translated into metabolically costly proteins,[87] which appears to be implicit in Dickinson's[17] comparative electrophoresis results discussed above. Heterologous transcription of isolated genes into expression vectors derails a significant amount of the host cell's resources, imposing a metabolic load in the case of unicellular prokaryotes.[88]

In animal species, arresting overall transcription and translation rates is a major contributor to energy savings in stress-induced depressed metabolic states[89] suggesting that (i) at least during hypometabolism, widespread ectopic expression is a burden to the cell; and (ii) ectopic expression may be subjected to overall downregulatory mechanisms, although global depression of transcription could as well be readily interpreted as a side-product of the downregulation of specific key loci. But ectopic expression might also seem wasteful if we take into account, for example, that several otherwise seemingly insignificant properties, such as genome-wide codon-usage biases, can result from the benefits of increasing the efficiency of resource utilization and allowing more rapid growth.[90] Analogously, features like pseudogenes or introns, considered 'junk' at the time of their discovery, are increasingly being known to have functions.[32,47,91]

Besides metabolic costs, the hypothesized high numbers of many different inappropriate transcripts in a specialized cell, with the potential to generate as many different inappropriate peptides, would diminish the specificity of the differentiated state.[92] This reasoning has been advanced in order to explain the quantal expansion of gene numbers that took place at the prokaryote–eukaryote and invertebrate–vertebrate boundaries, as a side effect of the evolution of global repressor mechanisms (the nuclear envelope and histones, and DNA methylation, respectively) of unscheduled transcription.[92]

All the above arguments make the marginal benefit hypothesis appealing, because by endowing ectopic expression with functionality, the question of its metabolic—or developmental—cost dissolves. Still, from an operational perspective, the marginal benefit hypothesis has the discomforting feature that the fitness contributions attributed to nonessential aspects of genes' transcriptional profiles are probably so small that they can prove to be undetectable by neutrality tests of evolutionary transcriptomics.[13]

Discussions of the evolutionary significance of organismal features should separate the question of what is the primary evolutionary reason for their existence from the question of what is their current use in the life of the organism. How functionless, but energetically cost-ineffective, attributes could have evolved may be understandable if one looks at them as parts of the integrated whole system. Accordingly, ectopic expression may have originated as a side effect of a novel regulatory strategy, if the whole new regulatory system turned out to be metabolically less costly than the one that it replaced. But cost-ineffective ectopic expression may as well have evolved passively by genetic drift, in response to long-term population-size reductions concomitant to changes in regulatory architecture.[93] If ectopic expression were a costly gratuitous by-product of the regulatory system, one would expect that genes with a high level and breadth of expression, i.e. housekeeping genes, would show a tendency to cluster apart from genes that are expressed in a tissue-specific fashion. Such a trend could be observed, for example, in the genomes of warm-blooded vertebrates, in which housekeeping genes tend to be located in the central, open chromatin of the interphase nuclei, whereas specific genes tend to be in the peripheral, more compact chromatin.[78,94,95] If the account just given is correct, it would be an interesting case of a novel feature, i.e. the regulatory system, making a fruitful use of available parts, i.e. the genomic architecture. The genomic architecture would evolve more efficiently by accommodating major aspects of the regulatory system.

It is expected that the discovery of widespread ectopic expression will have a profound impact on current views of the evolution of physiology and development. Co-option, the recruitment of pre-existing units, genes, organs, or other structures, for new functions, has long been assumed to play a key role in the evolution of developmental and physiological novelties.[6–8] Genes can be co-opted to new roles by changing their patterns of regulation, by changing the functions of the proteins that they encode, or both.[96] However, little is known about how co-option takes place mechanistically, and its underlying evolutionary forces.[58,96] Gene duplication, followed by divergence of the paralogs, has long been thought to be a necessary first step for the evolution of most novel functions.[97] But gene co-option can take place without duplication, via changes in parts of the amino acid sequence not required for the current function, or by regulatory changes that alter the expression profiles.[96] These last two mechanisms have received less attention, because of the belief that the evolution of a gene is generally strongly constrained owing to negative trade-offs associated with the necessity to maintain its original function.[96] This belief has recently been challenged by a laboratory-directed evolution study, showing that it is possible to endow an enzyme with novel, promiscuous functions without compromising its native function (a property called 'functional promiscuity evolvability'[98,99]). By pervasively exposing highly plastic proteins capable of rapidly evolving promiscuous functions, and the regulatory regions that control their expression to novel cellular environments different from those in which they evolved, ectopic expression may become a primary fosterer of functional innovation. Many genes may be responding to selective pressures for functional diversification well before duplications occur. After duplication, specialization of duplicate copies may allow for fine-tuning of these functions and independent patterns of expression. If the necessity to maintain the native function does not hamper the evolution of new functions, this would imply that the chances of gene neofunctionalization[100] are greater than previously assumed.[101,102] If some degree of functional promiscuity also holds for regulatory genes, including those encoding transcription factors and signaling proteins,[58,96] ectopic expression may facilitate their co-option for new functions and rewiring and/or waxing of the regulatory networks. Widespread ectopic expression, a gratuitous by-product of a novel regulatory system, may thus have been a primary motor for increases in genome structural and regulatory complexity in the evolution from prokaryotes to multicellular eukaryotes.

Conclusions

High rates of turnover of transcription-factor-binding sites together with the combinatorial control of functionally unrelated genes, concomitantly exposed to their regulatory effectors in open-chromatin state, suggest that ectopic expression might be a widespread phenomenon in the cells of multicellular eukaryotes. This inference, derived from molecular models, is consistent with the results of evolutionary transcriptomics studies, which indicate that many aspects of a gene transcriptional profile may have little if any adaptive significance. Altogether, the evidence herein discussed evinces that the received notion of the regulation of gene expression in multicellular eukaryotes as an exquisitely regulated process, as well as functionally meaningful, needs to be changed towards a view in which the regulation of functionally relevant expression produces much litter. It may be simplistic to assume that any transcriptional output that is not (yet) understood is noise. Regardless of its probable origin as a by-product of how gene expression is hardwired into the genome architecture, widespread ectopic expression has significant evolutionary potential. The results of protein electrophoretic assays[17–22] suggest that most ectopically

produced transcripts undergo translation. But ectopic expression could be functionally co-opted at other levels, as many as the decoding of genetic information can generate. The weak (if any) selective forces at play, together with the intricate ways that genetic information is encoded in multicellular eukaryotes, may handicap the discovery of functional ectopic expression. Functional ectopic expression may compensate for its metabolic drain on the cell, but functional prospects should be sought from the perspective of ectopic expression as an integral part of a phenotypic whole. Gene array methodologies are known to suffer from inaccuracies; they still need substantial refinement.[103] Future technical developments, together with more comprehensive evolutionary transcriptomics studies, involving denser taxon sampling and more closely related species, will provide increasingly powerful grounds for testing the strict neutrality model. The possibility of functional ectopic expression implies that all features of a gene transcriptional profile, including those that cannot be anticipated on the basis of known functions, deserve to be investigated.

Acknowledgments

We wish to express our gratitude to Adam S. Wilkins and Michael Lynch for their valuable comments, which have much contributed to improve our manuscript.

References

1. Emerson BM. 2003. Specificity of gene regulation. Cell 109:267–270.
2. Fisher RA. 1930. The Genetical Theory of Natural Selection. New York: Dover Press.
3. Haldane JBS. 1932. The Causes of Evolution. New York: Longman.
4. Schrödinger E. 1944. What is Life? Cambridge: Cambridge University Press.
5. Monod J. 1971. Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology. Translated from the French by Austryn Wainhouse. New York: Knopf.
6. Davidson EH. 2001. Genomic Regulatory Systems—Development and Evolution. San Diego: Academic Press.
7. Carroll SB, Grenier JK, Weatherbee SD. 2001. From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design. Malden, MA: Blackwell Scientific.
8. Wilkins AS. 2002. The Evolution of Developmental Pathways. Sunderland MA: Sinauer Associates.
9. Akam M. 1998. *Hox* genes, homeosis and the evolution of segment identity: no need for hopeless monsters. Int J Dev Biol 42:445–451.
10. García-Bellido A. 1975. Genetic control of wing disc development in *Drosophila*. In Cell Patterning. Ciba Found Symp 29:161–178.
11. Morata G, Lawrence PA. 1977. Homeotic genes, compartments and cell determination in *Drosophila*. Nature 265:211–216.
12. Yanai I, Graur D, Ophir R. 2004. Incongruent expression profiles between human and mouse orthologous genes suggest widspread neutral evolution of transcription control. OMICS 8:15–24.
13. Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, et al. 2004. A neutral model of transcriptome evolution. PloS Biol 2:682–689.
14. Rodríguez-Trelles F. 2004. Transcriptome evolution: much ado about nothing? Heredity 93:405–406.
15. Dickinson WJ. 1988. On the architecture of regulatory systems: evolutionary insights and implications. Bioessays 8:204–208.
16. Spellman PT, Rubin GM. 2002. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. BMC J Biol 1:5.1–5.8.
17. Dickinson WJ. 1980. Evolution of patterns of gene expression in Hawaiian Picture-winged *Drosophila*. J Mol Evol 16:73–94.
18. Powell JR, Rico M, Andjelkovic M. 1980. Population genetics of *Drosophila* amylase. Evolution 34:209–213.
19. Thorpe PA, Loye J, Rote CA, Dickinson WJ. 1993. Evolution of regulatory genes and patterns: relationships to evolutionary rates and to metabolic functions. J Mol Evol 37:590–599.
20. Morton RA, Singh RS. 1985. Biochemical properties, homology, and genetic variation of *Drosophila* 'nonspecific' esterases. Biochem Genet 23:959–973.
21. Ross J, Fong PP, Cavener DR. 1994. Coordinate evolution of the cis-acting regulatory elements and developmental expression of the *Drosophila Gld* gene in seven species from the melanogaster subgroup. Dev Genet 15:38–50.
22. Thorpe PA, Dickinson WJ. 1988. The use of regulatory patterns in constructing phylogenies. Syst Zool 37:97–105.
23. Bowers CW. 1994. Superfluous neurotransmitters? TINS 8:315–320.
24. Dickinson WJ, Sullivan DT. 1975. Gene-enzyme systems in Drosophila. New York: Springer.
25. Zuckerkandl E. 1978. Multilocus enzymes, gene regulation and genetic sufficiency. J Mol Evol 12:57–89.
26. Goebl MG, Petes TD. 1986. Most of the yeast genomic sequences are not essential for cell growth and division. Cell 46 :983–992.
27. Kwiatowski J, Skarecky D, Tyler R, Pascual M, Ayala FJ. 1991. A truncated P element is inserted in the transcribed region of the Cu,Zn SOD gene of an SOD ''null'' strain of *Drosophila melanogaster*. Free Rad Res Comms 12–13:429–435.
28. Rossant J, Moens C, Skarnes W, Gasca S, Hill D, et al. 1991. Identification and mutation of developmentally regulated genes in the mouse. FASEB J 5:A783.
29. Erickson HP. 1993. Gene knockouts of *c-src*, Transforming Growth Factor β1, and Tenascin suggest superfluous, nonfunctional expression of proteins. J Cell Biol 120:1079–1081.
30. Cook DL, Gerber AN, Tapscott SJ. 1998. Modelling stochastic gene expression: implications for haploinsufficiency. Proc Natl Acad Sci USA 95:15641–15646.
31. Veitia RA. 2002. Exploring the etiology of haploinsufficiency. Bioessays 24:175–184.
32. Balakirev ES, Ayala FJ. 2003. Pseudogenes: are they ''junk'' or functional DNA? Ann Rev Genet 37:123–151.
33. Boger ET, Sellers JR, Friedman TB. 2001. Human myosin XVBP is a transcribed pseudogene. J Muscle Res Cell Motil 22:447–483.
34. Edgar AJ. 2002. The human L-threonine 3-dehydrogenase gene is an expressed pseudogene. BMC Genet 3:18.1–18.13.
35. Lande R. 1976. Natural selection and random genetic drift in phenotype evolution. Evolution 30:314–334.
36. Lynch M, Hill WG. 1986. Phenotypic evolution by neutral mutation. Evolution 40:915–935.
37. Lynch M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. Am Nat 136:246–256.
38. Cheetam AH, Jackson JBC, Hayec LC. 1993. Quantitative genetics of bryozoan phenotypic evolution. I. Rates for random change versus selection in differentiation of living species. Evolution 47:1526–1538.
39. Felsenstein J. 2004. Inferring Phylogenies. Sinauer Associates, Sunderland, MS.
40. Oleksiak MF, Churchill GA, Crawford DL. 2002. Variation in gene expression within and among natural populations. Nat Genet 32:261–266.
41. Rifkin SA, Kim J, White KP. 2003. Evolution of gene expression in the *Drosophila melanogaster* subgroup. Nat Genet 33:138–144.
42. Uddin M, Wildman DE, Liu G, Xu W, Johnson RM, et al. 2004. Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. Proc Natl Acad Sci USA 101:2957–2962.
43. Felsenstein J. 1985. Phylogenies and the comparative method. Am Nat 125:1–15.
44. Arnason U, Gullberg A, Burgeuete AS, Janke A. 2000. Molecular estimates of primate divergences and new hypotheses for primate dispersal and the origin of modern humans. Hereditas 133:217–228.
45. Rodríguez-Trelles F, Tarrío R, Ayala FJ. 2002. A methodological bias toward overestimation of molecular evolutionary timescales. Proc Natl Acad Sci USA 99:8112–8115.

46. Graur D, Martin W. 2003. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. TIG 20:80–86.
47. Podlaha O, Zhang JZ. 2004. Nonneutral evolution of the transcribed pseudogene Makorin1-p1 in mice. Mol Biol Evol 21:2202–2209.
48. Hsieh W-P, Chu T-M, Wolfinger RD, Gibson G. 2003. Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. Genetics 165:747–757.
49. Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM. 2004. Common Pattern of Evolution of Gene Expression Level and Protein Sequence in *Drosophila*. Mol Biol Evol 21:1308–1317.
50. Enard W, Khaitovich P, Klose J, Zollner S, Heissing F, et al. 2002. Intra- and interspecific variation in primate gene expression patterns. Science 296:340–343.
51. Gu J, Gu X. 2003. Induced gene expression in human brain after the split from chimpanzee. TIG 19:63–65.
52. Karaman MW, Houck ML, Chemnick LG, Nagpal S, Chawannakul D, et al. 2003. Comparative analysis of gene-expression patterns in human and African great ape cultured fibroblasts. Genome Research 13:1619–1630.
53. Wray GA, Hahn MW, Abouheif A, Balhoff JP, Pizer M, et al. 2003. The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol 20:1377–1419.
54. Keightley PD, Lercher MJ, Eyre-Walker A. 2005. Evidence for widespread degradation of gene control regions in hominid genomes. Plos Biol 3:2e42.
55. Gu Z, Nicolae D, Lu HH-S, Li W-H. 2002. Rapid divergence in expression between duplicate genes inferred from microarray data. TIG 18:609–613.
56. Gu Z, Rifkin SA, White KP, Li W-H. 2004. Duplicate genes increase diversity within and between species. Nat Genet 36:577–579.
57. Huminiecki L, Wolfe KH. 2004. Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. Genome Res 14:1870–1879.
58. Duboule D, Wilkins AS. 1998. The evolution of 'bricolage'. TIG 14:54–58.
59. Stone JR, Wray GA. 2001. Rapid evolution of cis-regulatory sequences via local point mutations. Mol Biol Evol 18:1764–1770.
60. MacArthur S, Brookfield JFY. 2004. Expected rates and modes of evolution of enhancer sequences. Mol Biol Evol 21:1064–1073.
61. Rockman MV, Wray GA. 2002. Abundant raw material for cis-regulatory evolution in humans. Mol Biol Evol 19:1981–1990.
62. Tournamille C, Colin Y, Cartron JP, Kim CLV. 1995. Disruption of a GATA motif in the *Duffy* gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nat Genet 10:224–228.
63. Labie D, Pagnier J, Lapoumeroulie C, Roubahi F, Dunda-Belkhodia O, et al. 1985. Common haplotype dependency of high G gamma-globin gene expression and high Hb F levels in beta-thalassemia and sickle cell anemia patients. Proc Natl Acad Sci USA 82:2111–2114.
64. Lomvardas S, Thanos D. 2002. Modifying gene expression programs by altering core promoter chromatin architecture. Cell 110:261–271.
65. Kadam S, Emerson BM. 2002. Mechanisms of chromatin assembly and transcription. Curr Opin Cell Biol 14:262–268.
66. Cunliffe VT. 2003. Memory by modification: the influence of chromatin structure on gene expression during vertebrate development. Gene 305:141–150.
67. Gasser SM. 2001. Positions of potential: nuclear organization and gene expression. Cell 104:639–642.
68. Gierer A. 1973. Molecular models and combinatorial principles in cell differentiation and morphogenesis. Cold Spring Harbor Symp Quant Biol 38:951–961.
69. Kauffman SA. 1981. Pattern formation in the Drosophila embryo. Phil Trans Roy Soc Lond B 295:567–595.
70. Echols H. 1986. Multiple DNA-rpotein interactions governing high-precision DNA transactions. Science 233:1050–1056.
71. Ghazi A, VijayRaghavan K. 2000. Control by combinatorial codes. Nature 408:419–420.
72. Buchler N, Gerland U, Hwa T. 2003. On schemes of combinatorial transcription logic. Proc Natl Acad Sci USA 100:5136–5141.
73. Tsong AE, Miller MG, Raisner RM, Johnson AD. 2003. Evolution of a combinatorial transcriptional circuit: a case study in yeasts. Cell 115:389–399.
74. Travers A. 1993. DNA-Protein Interactions. London: Chapman and Hall.
75. Latchman DS. 1998. Eukaryotic Transcription Factors. San Diego: Academic Press.
76. Arnone MI, Davidson EH. 1997. The hardwiring of development: organization and function of genomic regulatory systems. Development 124:1851–1864.
77. Levine M, Tjian R. 2003. Transcription regulation and animal diversity. Nature 424:147–151.
78. De Laat W, Grosveld F. 2003. Spatial organization of gene expression: the active chromatin hyb. Chromosome Res 11:447–459.
79. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA. 2004. Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol 14:283–291.
80. Fourel G, Magdinier F, Gilson É. 2002. Insulator dynamics and the setting of chromatin domains. Bioessays 26:523–532.
81. Mattick JS. 2003. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. BioEssays 25:930–939.
82. Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, et al. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. Science 296:916–919.
83. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, et al. 2004. Unbiased Mapping of Transcription Factor Binding Sites Along Human Chromosomes 21 and 22 Points to Widespread Regulation of Non-Coding RNAs. Cell 116:499–511.
84. Kampa D, Cheng J, Kapranov P, Yamanaka M, Brubaker S, et al. 2004. Novel RNAs Identified from an In-depth Analysis of the Transcriptome of Human Chromosomes 21 and 22. Genome Res 14:331–342.
85. Johnson JM, Edwards S, Shoemaker D, Schadt EE. 2005. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. TIG 21:93–102.
86. Tatcher JW, Shaw JM, Dickinson WJ. 1998. Marginal fitness contributions of nonessential genes in yeast. Proc Natl Acad Sci USA 95:253–257.
87. Rolfe DFS, Brown GC. 1997. Cellular energy utilization and molecular origin of standard metabolic rate in mammals. Physiological Reviews 77:731–758.
88. Glick BR. 1995. Metabolic load and heterologous expression. Biotechnology Advances 13:247–261.
89. Storey KB, Storey JM. 2004. Metabolic rate depression in animals: transcriptional and translational controls. Biol Rev 79:207–233.
90. Akashi H. 2001. Gene expression and molecular evolution. Curr Opin Genet Dev 11:660–666.
91. Rodríguez-Trelles F, Tarrío R, Ayala FJ. 2003. Evolution of cis-regulatory regions versus codifying regions. Int J Dev Biol 47:665–673.
92. Bird AP. 1995. Gene number, noise reduction and biological complexity. TIG 11:94–100.
93. Lynch M, Conery JS. 2003. The origins of genome complexity. Science 302:1401–1404.
94. Saccone S, Federico C, Bernardi G. 2002. Localization of the gene-richest and the gene-poorest isochors in the interphase nuclei of mammals and birds. Gene 300:169–178.
95. Vinogradov AE. 2004. Compactness of human housekeeping genes: selection for economy or genomic design? TIG 20:248–253.
96. True JR, Carroll SB. 2002. Gene co-option in physiological and morphological evolution. Ann Rev Cell Dev 18:53–80.
97. Ohno S. 1970. Evolution by gene duplication. Springer-Verlag, NY.
98. Aharoni A, Gaidukov L, Khersonsky O, Gould SM, Roodveldt C, et al. 2005. The 'evolvability' of promiscuous protein functions. Nat Genet 37:73–76.
99. Kondrasov FA. 2005. In search of the limits of evolution. Nat Genet 37:9–10.
100. Rodríguez-Trelles F, Tarrío R, Ayala FJ. 2003. Convergent neofunctionalization by positive Darwinian selection after ancient recurrent duplications of the *xanthine dehydrogenase* gene. Proc Natl Acad Sci USA 100:13413–13417.
101. Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155.
102. Lynch M, O'Hely M, Walsh B, Force A. 2001. The probability of preservation of a newly arisen gene duplicate. Genetics 159:1789–1804.
103. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. 2004. Noise minimization in eukaryotic gene expression. PLoS Biol 2:1–2.