CpG content affects gene silencing in mice: evidence from novel transgenes

Christine Chevalier, Isabelle Henry, Lucile Montfort, Suzanne Capgras, Sylvie Forlani, John Muschler, Jean-François Nicolas

HAL Id: pasteur-03229787
https://pasteur.hal.science/pasteur-03229787

Submitted on 19 May 2021

Research

CpG content affects gene silencing in mice: evidence from novel transgenes

Christine Chevalier-Mariette¤*, Isabelle Henry¤*, Lucile Montfort¤†, Suzanne Capgras*, Sylvie Forlani* , John Muschler‡ and Jean-François Nicolas*

Addresses: *Unité de Biologie moléculaire du Développement, Institut Pasteur, rue du Dr Roux, 75724 Paris Cédex 15, France. †Current address: Vividis Biotech, rue de l'Espinay, Québec, QC, Canada G1L 3L5. ‡Current address: Life Science Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720, USA.

¤ These authors contributed equally to this work.

Correspondence: Jean-François Nicolas. E-mail: jfnicola@pasteur.fr

Abstract

Background: Transgenes are often engineered using regulatory elements from distantly related genomes. Although correct expression patterns are frequently achieved even in transgenic mice, inappropriate expression, especially with promoters of widely expressed genes, has been reported. DNA methylation has been implicated in the aberrant expression, but the mechanism by which the methylation of a CpG-rich sequence can perturb the functioning of a promoter is unknown.

Results: We describe a novel method for analyzing epigenetic controls that allows direct testing of CpGs involvement by using *LacZ* reporter genes with a CpG content varying from high to zero that are combined with a CpG island-containing promoter of a widely expressed gene - the α-subunit of the translation elongation factor 1. Our data revealed that a *LacZ* transgene with null CpG content abolished the strong transgene repression observed in the somatic tissues of transgenic lines with higher CpG content. Investigation of transgene expression and methylation patterns suggests that during *de novo* methylation of the genome the CpG island-containing promoter escapes methylation only when combined with the CpG-null transgene. In the other transgenic lines, methylation of the promoter may have led to transcriptional silencing.

Conclusions: We demonstrate that the density of CpG sequences in the transcribed regions of transgenes can have a causal role in repression of transcription. These results show that the mechanism by which CpG islands escape *de novo* methylation is sensitive to CpG density of adjacent sequences. These findings are of importance for the design of transgenes for controlled expression.

Background

Methylation of cytosine residues of the CpG dinucleotides of DNA constitute the basis of an epigenetic control of gene expression in vertebrate animals [1,2]. Genomic methylation patterns are of critical importance in various biological processes such as silencing of parasitic elements, development, tumorigenesis and genomic imprinting [3-5]. In the genome, CpGs are not uniformly distributed: the average level is only 1% and there are about 30,000 short regions rich in CpGs [6]. Frequently, the promoters of widely expressed genes are included in a CpG island whereas promoters of strictly expressed genes generally are not [7]. In mammals, DNA methylation patterns are established by two distinctive DNA cytosine methyltransferases, Dnmt3a and Dnmt3b, during development after implantation [8]. DNA methylation is then maintained by another methyltransferase, Dnmt1 [9]. CpG islands remain unmethylated; their aberrant methylation results in the silencing of their expression. Methylation of promoters does not lead to silenced transcription until chromatin proteins are recruited to the region [10]. Methyl cytosine binding proteins (MBPs), associated with 5meCpG and also part of complexes that contain histone deacetylases (HDACs), are involved in silencing. Dnmts also associate with HDAc as well as with HP1 [11]. The DNA of these silent regions is packaged into nucleosomes that contain deacetylated histone H3 [2,12].

Effects that spread in *cis* have been demonstrated at two levels in this process. *De novo* methylation can spread from regions that serve at foci [13]. Methylation of non-promoter sequences can result in transcriptional silencing of a reporter gene [10,14]. Diffusion of gene silencing involves histone deacetylation as well as structural and remodeling activities of chromatin [15]. The mechanisms underlying these effects in *cis* are not understood. If both methylation and the effects of methylation can diffuse in *cis*, the questions of how CpG islands escape methylation when adjacent sequences are hypermethylated, and how the resident promoter escapes silencing, are clearly posed. It could be that boundary elements delimit domains allowing them to maintain different states [16]. Alternatively, proteins such as CpG-binding proteins [17], which bind unmethylated DNA, are potential candidates for maintaining hypomethylation of CpG.

These issues are important for understanding the functional organization of the genomes. They are also important when designing artificial genes from sequences of distant origin. In fact it is remarkable that the CpG-rich bacterial sequences associated with eukaryotic control elements are functional. However, there are cases where expression is not faithfully reproduced. This is almost always the case with the association of bacterial sequences with promoters of widely expressed genes [18-22]. This could be due to the absence of required enhancers or Locus control region (LCR) elements or to their association with the bacterial sequences. In one case it

was shown that the rewriting of the sequences of a bacterial gene was beneficial for its expression in transgenic mice [23].

To evaluate the significance of global CpG content in epigenetic controls, we constructed *LacZ* genes differing only in their CpG content, from high density to null [24]. These molecules were combined with promoters of widely expressed genes - promoters that generally do not reproduce a widespread expression pattern when used in transgenesis. We show that a complete repression of the CpG-rich *LacZ* transgene is observed even at single copy in all somatic tissues whereas widespread expression is obtained with the *LacZ* transgene lacking CpG. Methylation studies suggest that the mechanism by which this effect is mediated involves interference between the protection of a CpG island from *de novo* methylation and the *de novo* methylation of the adjacent sequences during development.

Results

Two novel reporter genes: *LagZ* and *LagoZ*

Two novel reporter genes, *LagZ* and *LagoZ*, were derived by directed mutagenesis of the *LacZ* gene which has a CpG content of 9.24% (291 CpG for 3,076 bp, Figure 1a,b), which corresponds in density to CpG-rich regions of the vertebrate genome. *LagZ* (Figure 1c) and *LagoZ* (Figure 1d) have CpG densities of 1.6% (50% above CpG-poor regions in vertebrates) and 0.06% (2 CpG), respectively. Through use of alternate codons, each modified codon encodes the same amino acid as *LacZ* [24]. A few non-conservative mutations have appeared during the mutagenesis in *LagoZ* [24]. As we restricted the modification to CpG sequences, the AT content of the three DNA molecules remained almost the same (44.2%, 50.7% and 51.8%, respectively) and the differences in codon usage were minimal. We chose this approach, rather than a general re-encoding of *LacZ* with preferred mammalian codons, to test unambiguously the differences in transgene behavior resulting from changes in CpG content alone. These genes were attached to the same control elements, which included a 2.3 kb fragment of the human promoter of the α-subunit of the elongation factor 1 of translation (EF1α) [25]. This promoter possesses the characteristics for widely expressed genes in particular and it is included in a CpG island with a CpG content of 6.2% (Figure 1). The reporter gene was placed in exon 2 at 160 bp from the end of the CpG-rich island. We chose widely expressed genes as it is particularly with ubiquitous promoters that transgenes have failed to reproduce correct expression patterns, despite many attempts [18,19,21]. The expression potential of these three transgenes was analyzed *in vivo* through the creation of transgenic mice. The testing of these transgenes in animals (as opposed to cultured cells) permitted analysis of transgene expression at multiple periods of development and in multiple cell types: here, during gametogenesis and at the blastocyst stage, when the genome is hypomethylated, and after
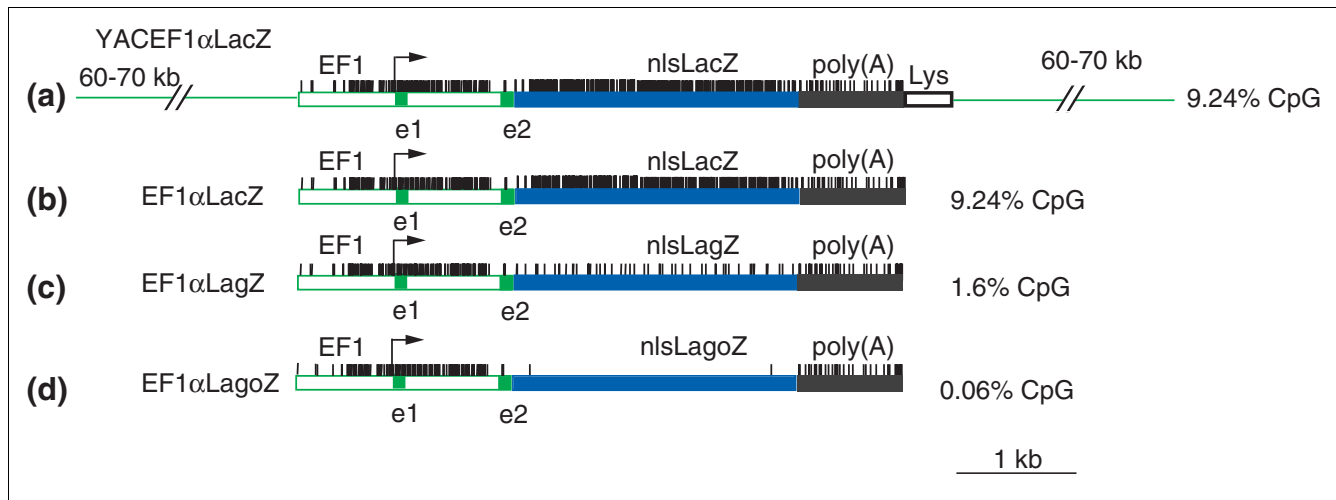
Figure 1
EF1αLacZ reporter genes with different CpG content. All four constructs share the same promoter region (EF1) from the human EF1α (translation elongation factor 1, α subunit) gene including the exon 1 (e1) and a part of exon 2 (e2), a polyadenylation sign (poly(A)) from Moloney murine virus and a nuclear locating signal (nls) from the large T antigen of SV40 [32]. Both the left and right arms of the YAC (green line) are 60-70 kb long. Each vertical bar corresponds to a CpG dinucleotide. The percentage of CpGs in the reporter gene is indicated on the right.

embryonic day 9.5 (E9.5), when the genome of somatic cells is hypermethylated [26,27].

Expression studies of EF1αLacZ, EF1αLagZ and EF1αLagoZ in transgenic animals
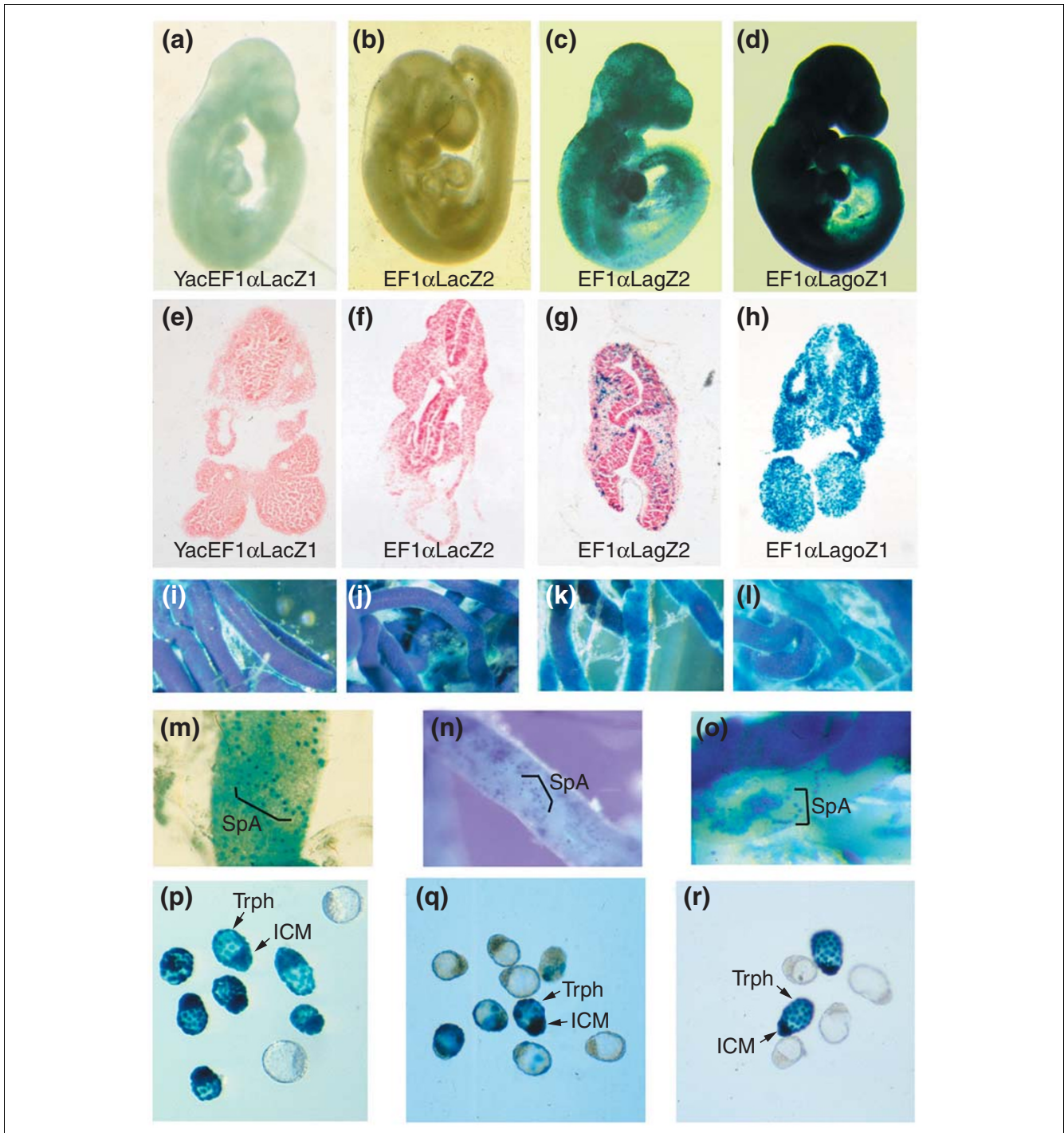
In all four *LacZ* (Figure 2i,j,m), four *LagZ* (Figure 2k,n) and four *LagoZ* (Figure 2l,o) transgenic lines, the male germ cells expressed the transgenes. Expression includes type A spermatogonia (SpA), the small chains in Figure 2m-o. Therefore, the alteration of CpG density in the *LagZ* and *LagoZ* sequences had no gross consequence on expression in these cells. The quantitative expression of the three transgenes microinjected in 1-cell stage embryo was also indistinguishable (5.6, 5.7 and 7.1 × 10⁻⁵ U β-gal, respectively, for *LacZ*, *LagZ* and *LagoZ*). Together, these observations suggest comparable RNA stability and minimal effects of codon usage on gene expression between the three molecules.

In contrast, dramatic differences were observed in the embryo between the three transgenes (see Figure 2a-d for representative examples, and Table 1). A systematic, high expression of the *LagoZ* transgene was always observed in embryonic and extra-embryonic tissues whatever the line (n = 4) (Figure 2d,h), whereas no expression of the *LacZ* transgene was observed (n = 4) (Figure 2b,f) and only a variegated expression was observed in one *LagZ* line (n = 4) (EFIαLagZ2, Figure 2c,g). Following differentiation of cell types, for example at P7, these differences in transgene expression were maintained (Table 1). Clearly, epigenetic controls are imposed early on the *LacZ* and *LagZ* transgenes, and cell differentiation does not erase these controls.

To test when this epigenetic control is imposed on the *LacZ* and *LagZ* transgenes, β-galactosidase (β-gal) expression was searched for in blastocysts. Both *LacZ* and *LagZ* transgenes are strongly expressed in the ICM (inner cell mass) and trophectoderm (Figure 2q,r) in the four EF1αLac and four EF1αLag transgenic lines. Therefore, the epigenetic control is imposed on the genome after the blastocyst stage but before E9.5. Altogether, these results show that the high density of CpGs in the transcribed region of *LacZ* is the cause of total repression of its expression in somatic cells. A low density of CpGs (*LagZ* gene), although higher than the corresponding sequences in EF1α gene, still provoked a total repression or, at best, a variegated expression in tissues included in the widespread expression pattern of the EF1α promoter. An absence of CpG sequences in the transcribed part of the gene (*LagoZ*) resulted in a complete release from this repression.

Expression studies of single-copy YacEF1αLacZ transgenic animals

Some epigenetic controls are especially effective on repeated sequences in the genome [28]. However, the study of four YacEF1αLacZ transgenic mice (Figure 1a) in which the transgene is at single copy indicated that, even in this condition, the CpG content of the transcribed region caused total repression. The expression pattern is indistinguishable from the expression pattern of EF1αLacZ: no expression of the *LacZ* transgene in embryonic and extra-embryonic tissues and no expression at P7 (Table 1, n = 5). As for EF1αLacZ, the male germ cells and the cells at the blastocyst stage (Figure 2p) expressed the transgene.

Figure 2

Reduced CpG content in *LacZ* abolishes gene silencing in somatic tissues. **(a-h)** YacEF1αLacZ, EF1αLacZ, EF1αLagZ and EF1αLagoZ transgene expression in E9.5 representative postimplantation embryos: (a-d) *in toto* X-gal stained E9.5 embryos, (e-h) 20 μm cryostat sections at the level of the anterior CNS. (a,e) YacEF1αLacZ1, (b,f) EF1αLacZ2, (c,g) EF1αLagZ2 and (d,h) EF1αLagoZ1 transgenic lines. **(i-o)** Transgene expression in testis, *in toto* X-gal staining of the seminiferous tubules: (i-l) general view, (m-o) sectors with type A spermatogonia. (i) YacEF1αLacZ1, (j,m) EF1αLacZ2, (k,n) EF1αLagZ2 and (l,o) EF1αLagoZ1 transgenic lines. (m-o) β-gal+ type A spermatogonia (SpA). **(p-r)** Expression patterns at the blastocyst stage: (p) YacEF1αLacZ1, (q) EF1αLacZ2 and (r) EF1αLagZ2 blastocysts X-gal stained. Trph, trophectoderm; ICM, inner cell mass.

Table 1

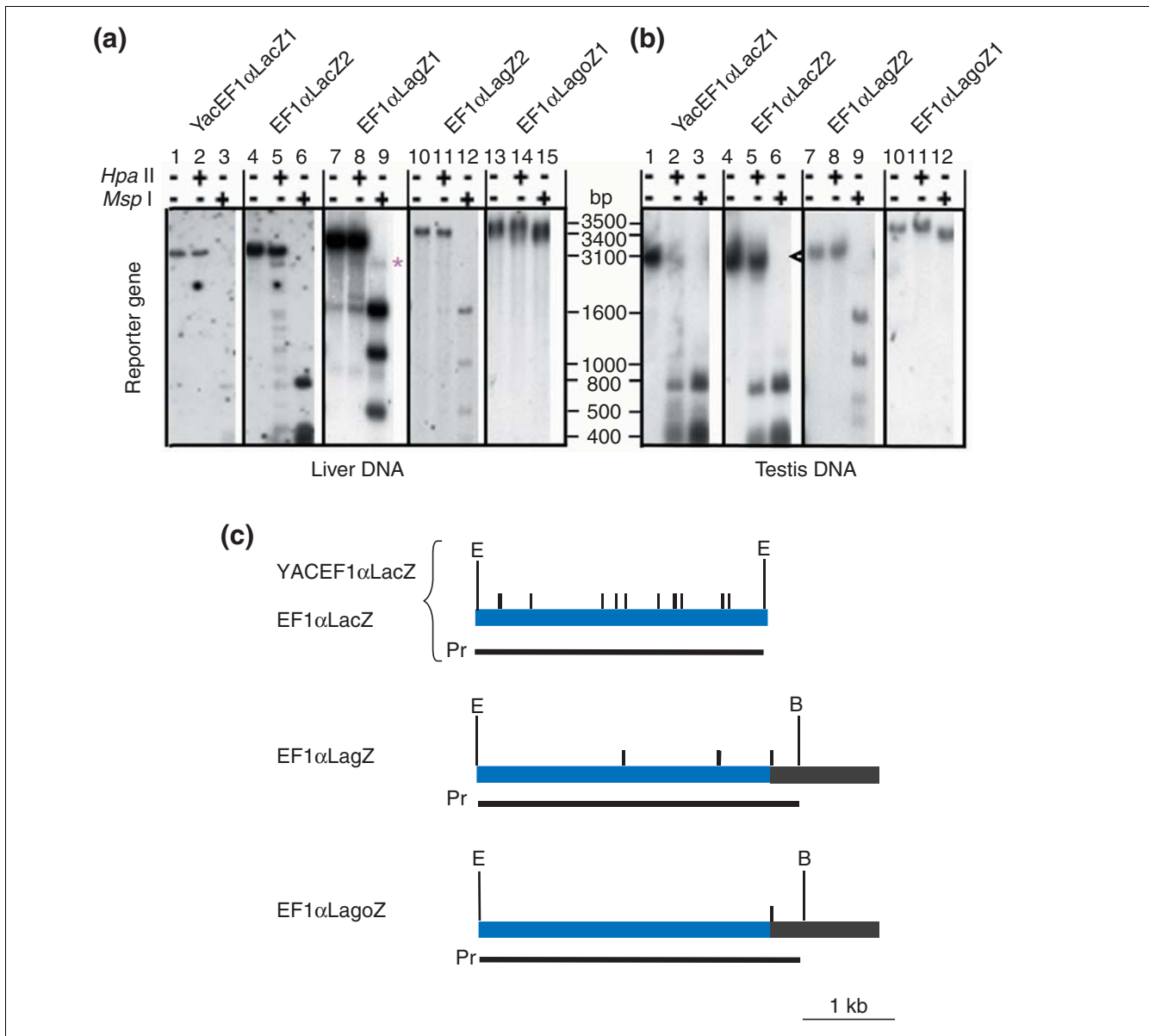**Summary of LacZ expression pattern in transgenic animals**

		YAC EF1αLacZ					EF1αLacZ				EF1αLagZ				EF1αLagoZ			
Line (number of copies)		1 (1)	2 (1)	3 (1)	4 (2)	5 (1)	1 (N)	2 (6)	3 (N)	4 (N)	1 (>6)	2 (2)	3 (5)	4 (N)	1 (2)	2 (10-20)	3 (N)	4 (10-20)
Germ cells (pachytene)	Male	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Female	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Mid-gestation embryo (E9.5)		-	-	-	-	-	-	PD	PD	-	+	+	PD	PD	+	+	+	+
P7	CNS	-	-	-	-	-	-	-	-	-	+	+	+	-	+	+	+	+
	Epidermis	-	-	-	-	-	-	-	-	-	+	+	-	-	+	+	+	+
	Heart auricle	-	-	-	-	-	-	-	-	-	+	+	-	-	+	+	+	+
	Heart	-	-	-	-	-	-	-	-	-	+	+	+	-	+	+	+	+
	Endothelial cells	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	+
	Liver	-	-	-	-	-	-	-	-	-	+	+	-	-	+	+	+	+
	Lung	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	+
	Spleen	-	-	-	-	-	-	-	-	-	+	+	-	-	+	+	+	+
	Thymus	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	+
	Skeletal muscle	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	+
	Tendon cells	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	+
Adults	CNS	-	-	-	-	-	-	-	-	-	+	+	+	-	+	+	+	+
	Epidermis	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	+
	Heart auricle	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	+
	Heart	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-
	Endothelial cells	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	+
	Liver	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	+
	Lung	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	-
	Spleen	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	+
	Skeletal muscle	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Tendon cells	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	-	+

+, β-galactosidase positive; -, β-galactosidase negative; N, number of copies >20; PD, position (integration site) dependent pattern of expression. Expression pattern is variegated in EF1αLagZ2 line.

The methylation patterns of *LacZ, LagZ* and of the promoter-containing EF1α sequences

This repression by CpG sequences is likely to be due to their methylation, but there are several possible hypotheses to explain the role of methylation. For instance, the methylation of the reporter gene alone can by itself provoke a change in chromatin structure of the adjacent promoter leading to its silencing, an idea compatible with the fact that the promoter included in a CpG island must escape methylation. Alternatively, direct methylation of the promoter can be necessary to repress the transgenes. To test these hypotheses, the methylation patterns of EF1αLacZ and EF1αLagZ, following digestion of DNA by *Msp*I or *Hpa*II, were analyzed in the liver, skin and brain as these tissues exhibit a high β-gal+ activity in the

EF1αLagoZ line (Table 1). We present the methylation patterns of the lines harboring the lowest copy number. Indeed, the presence of multiple copies in the other lines makes it impossible to correlate expression and methylation. Whatever the tissue or the line analyzed, all *Hpa*II sites (13 for *LacZ*, four for *LagZ*) were found to be methylated in 90% to 100% of the *LacZ* and *LagZ* fragments of β-gal- YacEF1αLacZ1, EF1αLacZ2 or EF1αLagZ1 lines (Figure 3a, lanes 2, 5 and 8, and data not shown, *n* = 5). Therefore, it is clear that the CpG-rich *LacZ* sequences are not recognized by the cells as CpG-rich islands as they are not protected from *de novo* methylation. Although these observations are compatible with the idea of an indirect repression of the reporter gene by methylated CpG, examination of the DNA of the β-gal+

Figure 3
Methylation of reporter sequences of the genomic DNA extracted from tissue of animals hemizygous for the transgene. DNA was extracted from **(a)** liver and **(b)** testis. **(c)** DNA was digested with the enzyme indicated under the blots: *Eco*RI (E) for YacEF1αLacZ and EF1αLacZ liberating a 3.1-kb-long fragment; *Eco*RI and *Bam*HI (B) for EF1αLagZ liberating a 3.4-kb-long fragment and EF1αLagoZ liberating a 3.5-kb-long fragment. Each vertical bar corresponds to an *Hpa*II/*Msp*I site. Probes are the complete reporter gene (Pr). For EF1αLagZ the *Eco*RI-*Bam*HI-*Msp*I digestion liberates one rearranged fragment indicated by a purple asterisk. The reference of the transgenic line is indicated at the top.

EF1αLagZ2 mice indicated that, at least in this case, the repressive effect could not be attributed solely to methylation of the reporter sequences. Indeed, in these β-gal+ mice, the *LagZ* gene was fully (100%) methylated (Figure 3a, lane 11).

If the methylation of the CpGs of the reporter gene is not sufficient to repress the transgene, is the methylation of the promoter sequences contained in a CpG-rich island involved? To address this issue we examined the two *Hpa*II DNA fragments specific to this region: the 166 bp fragment and the 167

bp fragment, and the combined 333 (166 + 167) bp EF1α fragments, the latter indicating partial methylation (Figure 4c). Surprisingly, we observed complete methylation of these sequences in β-gal- YacEF1αLacZ1 tissues (Figure 4a, lane 1, and data not shown, *n* = 3, the absence of both the 170 bp and the 333-bp-long fragments) and low levels of methylation in β-gal- EF1αLacZ2 and EF1αLagZ1 tissues (Figure 4a, lanes 3 and 5, *n* = 2). Therefore, methylation of the EF1α promoter sequences could explain the β-gal- phenotype. Other observations reinforced this possibility: in the variegated β-gal+
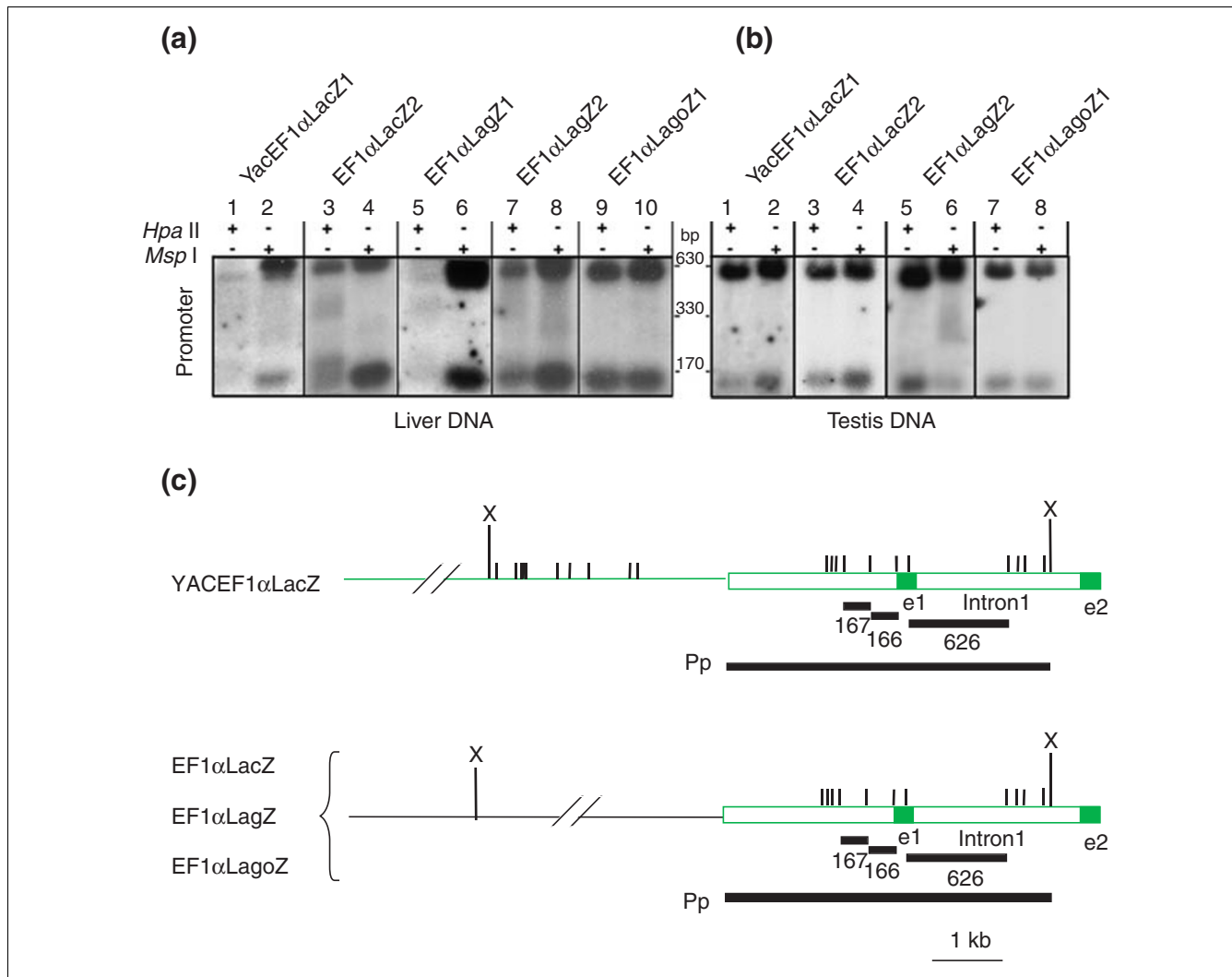
Figure 4
Correlation between an absence of methylation within the EFIα promoter and expression. Methylation of EF1α sequences of the genomic DNA extracted from **(a)** liver and from **(b)** testis of animals hemizygous for the transgene. DNA was digested with *Xho*I, which recognizes one site in the transgene and one site in the integration locus (X) (except for the YacEF1αLacZ, for which the two *Xho*I are in the transgene) and with *Hpa*II or *Msp*I as indicated at the top. The *Xho*I-*Msp*I digestion liberates a 626-bp-long fragment indicative of sites within the first intron and exon of EF1α and two 166 and 167-bp-long fragments indicative of sites within the promoter region. The reference of the transgenic line is indicated at the top. **(c)** All four EF1α reporter genes constructs share the same promoter region and the exon 1 (e1), the intron 1 and a part of exon 2 (e2) of EF1α. Each vertical bar corresponds to an *Hpa*II (*Msp*I) recognition site. Pp, EF1α probe.

EF1αLagZ2 mice, EF1α promoter sequences are only partially methylated, in contrast with β-gal⁻ EF1αLagZ1 (Figure 4a, lanes 7 and 5: the 170-bp-long band); in EF1αLagoZ1 mice, EF1α promoter sequences are not methylated at all (Figure 4a, lane 9: the 170-bp-long fragment).

The examination of other *Hpa*II sites in the CpG island demonstrated additional differences between the EF1αLagoZ1 line and the *LagZ* and *LacZ* lines. The 626-bp-long fragment in Figure 4a is indicative of the methylation of sites in the first exon and in the first intron of EF1α (Figure 4c; the 5'*Hpa*II fragments in EF1αLacZ2, EF1αLagZ1, EF1αLagZ2 and EF1αLagoZ1 transgenic lines are longer than 626 bp). This

fragment is fully detected in the EF1αLagoZ1 line (Figure 4a, lane 9) but only partially present or absent in *LagZ* and *LacZ* lines (Figure 4a, lanes 1, 3, 5 and 7, and data not shown, *n* = 6). Therefore, although the CpG island remains fully protected from methylation when combined with a *LagoZ* transgene (depleted of CpGs) it is only partially protected (at least when at low copy number, as in EF1αLagoZ1) or not protected at all when it is combined with CpG-containing sequences.

These results show firstly that β-gal⁻ phenotypes correlate with the methylation of EF1α promoter sequences (*n* = 5), and β-gal⁺ phenotypes with an absence of methylation of these sequences (*n* = 2); and secondly that the methylation of

the EF1α CpG-rich sequences is not observed when the reporter gene can not be methylated (EF1αLago Z1 line), indicating that, in this case, the CpG-rich island is protected from methylation. Thirdly, the results show that partial or complete methylation of this CpG island occurs when it is combined with CpG-rich sequences (*LacZ* or *LagZ*, *n* = 5) suggesting that, in this case, the CpG-rich island is not completely protected from methylation.

Expression and methylation patterns of YacEF1αLacZ, EF1αLacZ and EF1αLagZ in the male germ line

The expression patterns of YacEF1αLacZ, EF1αLacZ and EF1αLagZ suggest that inappropriate methylation of EF1α occurs during the period of *de novo* methylation of the genome after implantation of the embryo but before E9.5. Indeed, in all three cases the blastocysts (E4.5) of these lines strongly express the transgene (*n* = 13, Figure 2p-r) but the embryos at E9.5 and subsequent stages do not (Figure 2a-d, *n* = 12 out of 13). Two other observations support this conclusion. Firstly, the absence of methylation of the EF1α CpG island in YacEF1αLacZ1, EF1αLacZ2 and EF1αLagZ2 DNA in the male germ line (Figure 4b, lanes 1, 3 and 5: the presence of the 170 and 626-bp-long fragments, and data not shown, *n* = 5) and also, as expected, in EF1αLago Z1 (Figure 4b, lane 7). Secondly, the lower methylation level of the *LacZ* reporter sequences in this tissue when compared to somatic tissues (Figure 3b, lanes 2 and 5).

These observations also confirm that, as in the somatic tissues of EF1αLagZ2, a methylated *LagZ* reporter gene can correspond to a β-gal⁺ tissue (Figure 3b, lane 8). This situation also applies to *LacZ* reporter genes, as a significant fraction of the male germ cells harbor completely methylated *LacZ* reporters in EF1αLacZ2 (Figure 3b, lane 5, indicated by the arrowhead). Clearly, as in somatic tissues in the male germ line, the mere methylation of *LagZ* or *LacZ* is not sufficient to repress EF1α.

Discussion

We demonstrated that the density of CpG sequences in the transcribed regions of transgenes can have a causal role in the repression of transcription. The threshold density of CpG sequences required to initiate repression has not yet been determined, but repression is evident even when the CpGs are dispersed in the sequence and at a density just above that observed in the vertebrate genome (the EF1αLagZ lines). Therefore, we speculate that the distribution of CpG sequences within endogenous genes is adjusted, at least in part, as an adaptation to this potential repression.

Our results also clearly indicate that the CpG-rich *LacZ* sequences are not recognized as a CpG-rich island as they become hypermethylated (*n* = 5). Therefore, mere CpG content of a sequence is not sufficient to signal it as a CpG island. The additional *cis* signal(s) and *trans* factor(s) involved have

yet to be determined. Although still hypothetical, our results raise the possibility that among the 30,000 CpG-rich sequences of the mouse genome some may not be true CpG islands and therefore may influence expression of adjacent genes.

One way of explaining the repression of transgene expression by the methylation of CpGs is to postulate a global change in chromatin structure spreading in the unmethylated promoter [1,11,15]. This mechanism could be part of the repression observed with *LacZ* in somatic tissues as these sequences are hypermethylated, and since we did not observe repression without methylation of this sequence. However, this does not explain why certain *LagZ* lines express the transgene in somatic tissues even though *LagZ* is fully methylated, and why EF1αLagZ and EF1αLacZ mice express the transgene in the germ line even though *LagZ* and *LacZ* are also fully methylated in these cases.

An important observation made in this study suggests another hypothesis. The EF1α DNA fragment contains a CpG-rich island (in which the promoter region is included) which is expected to escape *de novo* methylation. We found that this is indeed the case when combined with *LagoZ* sequences that cannot be methylated (the EF1αLago Z1 line) but that the protection is less perfect when combined with ones that can be methylated. In this case, the EF1α DNA fragment is found to be methylated; therefore, the island is imperfectly protected from methylation (*n* = 6). According to these facts, a simple hypothesis for the mechanism of repression of transgenes by CpG-rich sequences can be proposed: these sequences could provoke the methylation of the adjacent sequences, which consequently could adopt a compact chromatin structure incompatible with transcription. It is interesting to note that, nevertheless, the relative protection of the EF1αLacZ sequences and the total protection of EF1αLago Z1 indicate that the *cis* element(s) which signal the EF1α sequences as a CpG-rich island are present in the 2.3 kb sequences used in our study, and that they function in many integration sites.

Analysis of the expression patterns of *LacZ* and *LagZ* genes then suggest that this inappropriate methylation of the CpG island occurs during the period of *de novo* methylation of the genome, shortly after implantation of the embryo [8]. Indeed the *LacZ* transgenes, EF1αLacZ and YacEF1αLacZ, are expressed in blastocysts and are completely repressed after implantation at E9.5 (*n* = 12). As it seems unlikely that sequences essential for protection against methylation are missing in the 2.3 kb EF1α fragment, we suggest that there is probably interference between the mechanism of protection of the CpG-rich island and the *de novo* methylation of *LacZ*. It should be noted that this interference does not correspond to an 'all-or-none' phenomenon as the individual EF1α DNA sequences are diversely methylated, even in the same mice.
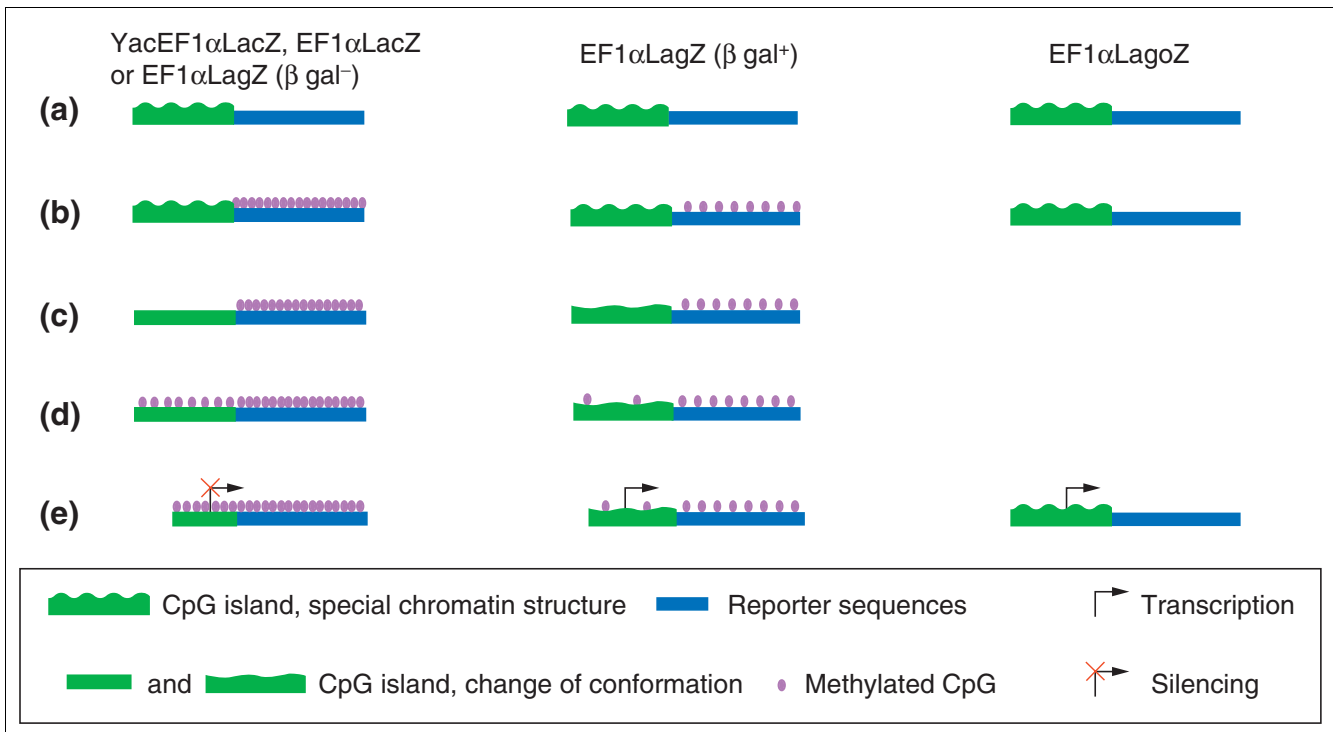
Figure 5
Model for the molecular mechanism of gene silencing mediated by DNA methylation of transcribed regions of a gene. **(a)** Before *de novo* methylation (at the blastocyst stage), EF1α and (maybe) the reporter gene are unmethylated. **(b)** During *de novo* methylation by the Dnmts, the reporter gene is methylated as it does not correspond to a protectable CpG-rich island. In contrast, the EF1α CpG-rich island is protected from methylation as it is in a special chromatin structure (the waves). **(c)** The partners of Dnmts and MBD induce a change of chromatin structure in EF1α (the modified waves). This may be involving methylated histones and/or HP1. **(d)** This change of chromatin structure interferes in part with the protection of the EF1α CpG-rich against CpG methylation. The degree of the interference is in relation with the density of methylation of the reporter gene (compare *LacZ* and *LagZ* [β-gal⁺]). **(e)** Chromatin conformation and transcription of the transgene in somatic tissue.

It is possible to speculate on the sequence of events leading to the methylation of EF1α. We propose that at the time of *de novo* methylation of the DNA, the CpG island is in a special chromatin structure, which may be due to fixation of proteins such as CpG-binding proteins which protect from methylation [17] (Figure 5a). The *de novo* methylation of the CpG-rich adjacent sequences (Figure 5b) will then lead to a change of conformation of the EF1α region (Figure 5c), which in turn will lead to the methylation of a few of its CpGs (Figure 5d). In the absence of this change of conformation (that is, the EF1αLagoZ case) the EF1α sequences will remain unmethylated. Alternatively, the *de novo* methylation of *LacZ* may spread in the adjacent CpG-rich island because of specific properties of the methylation complex. The EF1αLagZ (β-gal⁺) represents an intermediary situation suggesting a dose effect. Finally, in relation with the level of methylation of the EF1α sequences, the transgene will adopt a more or less compact chromatin structure allowing transcription or silencing the transgene (Figure 5e). With the CpG-free *LagoZ* sequences, the EF1α sequences could escape methylation, like the endogenous gene, and the transgene could adopt an open chromatin structure.

Apart from improving our understanding of transcriptional regulation *in vivo*, these results are of importance for the design of transgenes for controlled expression. CpG density is shown to be an important factor for the normal action of the mechanism through which CpG islands escape methylation. It is likely that many engineered genes are not adjusted to an optimal CpG density and are subjected to partial or total repression and/or are very susceptible to inactivation over time. To avoid these potential effects of CpG sequences, we propose the removal of all CpGs of the transcribed part of transgenes (employing alternate codons), as we have demonstrated that the absence of this dinucleotide largely eliminates repressive effects. Other experiments are in progress to determine whether the repressive effect of CpGs can be generalized to certain categories of genes with restricted patterns of expression like, for example, those with a CpG island located in their transcription start (61% of tissue-restricted genes) [7].

Conclusions
To evaluate the significance of global CpG content in epigenetic controls, we constructed *LacZ* genes differing only in

their CpG content, from high density to null [24]. These molecules were combined with promoters of widely expressed genes; promoters which generally do not reproduce a widespread expression pattern when used in transgenesis. We showed that a complete repression of the CpG-rich *LacZ* transgene is observed even at single copy in all somatic tissues whereas widespread expression is obtained with the *LacZ* transgene lacking CpG (the *LagoZ* construct). Methylation studies suggest that the mechanism by which this effect is mediated involves interference between the protection of a CpG island from *de novo* methylation and the *de novo* methylation of the adjacent sequences.

Materials and methods
Directed mutagenesis of *LacZ*
The CpG content of the *LacZ* plasmid was first lowered from 9.24% to 1.65% to generate the PytknlsLagZ plasmid and PytknlsLagZ was used as a template to lower the CpG content to 0.06% to generate PytknlsLagoZ. In Pytk plasmids, a bacterial promoter allowed an expression screening of the mutagenized genes in *LacZ*-defective bacteria (Δ-Lac). To construct *LagZ* and *LagoZ*, a polymerase chain reaction (PCR) technique was used in which the mutagenic oligonucleotide primers spanned the whole sequences but contained fewer CpG dinucleotides. These mutagenic primers were designed to preserve the integrity of the amino-acid sequence of the β-galatosidase reporter protein. After each run of PCR the DNA was purified, circularized and transfected in Δ-Lac bacteria to select the clones presenting β-galactosidase activity. The DNA sequences of *LagZ* and *LagoZ* were verified by sequencing [24].

The three versions of *LacZ* were associated to the promoter region of the human EF1α gene to generate pEFnlsLacZΔenh, pEFnlsLagZΔenh and pEFnlsLagoZΔenh plasmids. The *LacZ* reporter was also inserted by homologous recombination in yeast at the same location into a 150 kb YAC spanning the human EFIα locus (Figure 1a). Construction of plasmids and YAC will be described elsewhere. The expression of the different transgenes was controlled by transient expression experiments that gave the same activity [24] as measured using MUG quantification [29].

Transgenesis, recovery of embryonic and adult organs and cryosectioning
Each EF1α plasmid construct was first digested by the *Not*I restriction enzyme and then partially digested using the *Xho*I restriction enzyme to remove vector DNA sequences. The 7.9 kb *Xho*I/*Not*I inserts were purified and microinjected (200 copies/pl) into the male pronucleus one-cell embryo obtained from (C57BL/6JxDBA2) F1 superovulated females mated with males of the same strain. Five independent YacEF1αLacZ, four independent EF1αLacZ, four independent EF1αLagZ and four independent EIFαLagoZ mouse lines

were analyzed in this study. All embryos and organs were X-gal stained as described [30].

Real-time PCR on PE Biosystems Prism 7700
A TaqMan probe and primer set designed to amplify the *LacZ*, *LagZ* and *LagoZ* transgenes and an internal control (*Rapsyn* gene) were used to determine transgene copy number using DNA extracted from liver.

DNA preparation and methylation analysis
Tissues were collected from postnatal (P5) hemizygous mice (liver, skin and brain) and from adult males (testis) using standard protocol. Twenty micrograms of DNA were digested with appropriate enzymes to liberate the fragments to be analyzed - either *Hpa*II (methylation-sensitive) or *Msp*I (methylation-insensitive). The DNA was analyzed by Southern hybridization as previously described [31], probes for methylation are indicated in Figures 3 and 4.

References
1. Jones PA, Takai D: **The role of DNA methylation in mammalian epigenetics.** *Science* 2001, **293:**1068-1070.
2. Bird A: **DNA methylation patterns and epigenetic memory.** *Genes Dev* 2002, **16:**6-21.
3. Yoder JA, Walsh CP, Bestor TH: **Cytosine methylation and the ecology of intragenomic parasites.** *Trends Genet* 1997, **13:**335-340.
4. Li E, Beard C, Jaenisch R: **Role for DNA methylation in genomic imprinting.** *Nature* 1993, **366:**362-365.
5. Jones PA, Laird PW: **Cancer epigenetics comes of age.** *Nat Genet* 1999, **21:**163-167.
6. Takai D, Jones PA: **Comprehensive analysis of CpG islands in human chromosomes 21 and 22.** *Proc Natl Acad Sci USA* 2002, **99:**3740-3745.
7. Gardiner-Garden M, Frommer M: **CpG islands in vertebrate genomes.** *J Mol Biol* 1987, **196:**261-282.
8. Reik W, Walter J: **Genomic imprinting: parental influence on the genome.** *Nat Rev Genet* 2001, **2:**21-32.
9. Okano M, Bell DW, Haber DA, Li E: **DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.** *Cell* 1999, **99:**247-257.
10. Kass SU, Landsberger N, Wolffe AP: **DNA methylation directs a time-dependent repression of transcription initiation.** *Curr Biol* 1997, **7:**157-165.
11. Burgers WA, Fuks F, Kouzarides T: **DNA methyltransferases get connected to chromatin.** *Trends Genet* 2002, **18:**275-277.
12. Jones PA, Baylin SB: **The fundamental role of epigenetic events in cancer.** *Nat Rev Genet* 2002, **3:**415-428.
13. Jahner D, Jaenisch R: **Retrovirus-induced de novo methylation of flanking host sequences correlates with gene inactivity.** *Nature* 1985, **315:**594-597.
14. Hsieh CL: **Dependence of transcriptional repression on CpG methylation density.** *Mol Cell Biol* 1994, **14:**5487-5494.
15. Curradi M, Izzo A, Badaracco G, Landsberger N: **Molecular mechanisms of gene silencing mediated by DNA methylation.** *Mol Cell Biol* 2002, **22:**3157-3173.
16. Sun FL, Elgin SC: **Putting boundaries on silence.** *Cell* 1999, **99:**459-462.

17. Carlone DL, Skalnik DG: **CpG binding protein is crucial for early embryonic development.** *Mol Cell Biol* 2001, **21:**7601-7606.
18. Beddington RS, Morgernstern J, Land H, Hogan A: **An in situ transgenic enzyme marker for the midgestation mouse embryo and the visualization of inner cell mass clones during early organogenesis.** *Development* 1989, **106:**37-46.
19. Bonnerot C, Grimber G, Briand P, Nicolas JF: **Patterns of expression of position-dependent integrated transgenes in mouse embryo.** *Proc Natl Acad Sci USA* 1990, **87:**6331-6335.
20. Donoghue MJ, Morris-Valero R, Johnson YR, Merlie JP, Sanes JR: **Mammalian muscle cells bear a cell-autonomous, heritable memory of their rostrocaudal position.** *Cell* 1992, **69:**67-77.
21. Mehtali M, LeMeur M, Lathe R: **The methylation-free status of a housekeeping transgene is lost at high copy number.** *Gene* 1990, **91:**179-184.
22. Weis J, Fine SM, David C, Savarirayan S, Sanes JR: **Integration site-dependent expression of a transgene reveals specialized features of cells associated with neuromuscular junctions.** *J Cell Biol* 1991, **113:**1385-1397.
23. Scrable H, Stambrook PJ: **Activation of the lac repressor in the transgenic mouse.** *Genetics* 1997, **147:**297-304.
24. Henry I, Forlani S, Vaillant S, Muschler J, Choulika A, Nicolas JF: **LagoZ and LagZ, 2 genes depleted of CpG dinucleotides, derived from the LacZ gene for the study of epigenetic control.** *C R Acad Sci III* 1999, **322:**1061-1070.
25. Kim DW, Uetsuki T, Kaziro Y, Yamaguchi N, Sugano S: **Use of the human elongation factor 1 alpha promoter as a versatile and efficient expression system.** *Gene* 1990, **91:**217-223.
26. Monk M, Boubelik M, Lehnert S: **Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development.** *Development* 1987, **99:**371-382.
27. Reik W, Dean W, Walter J: **Epigenetic reprogramming in mammalian development.** *Science* 2001, **293:**1089-1093.
28. Lau S, Jardine K, McBurney MW: **DNA methylation pattern of a tandemly repeated LacZ transgene indicates that mast copies are silent.** *Dev Dyn* 1999, **215:**126-138.
29. Forlani S, Nicolas JF: **Quantification of minute levels of β-galactosidase. The example of individual 2 to 16-cell stage mouse embryos.** *Trends Genet* 1996, **12:**498-500.
30. Bonnerot C, Nicolas J-F: **Application of LacZ gene fusions to post-implantation development.** In: *Methods in Enzymology: Guide to Techniques in Mouse Development Volume 225*. Edited by: Wassarman PM, DePamphilis ML.San. Diego, CA: Academic Press; 1993: 451-469.
31. Bonnerot C, Legouy E, Choulika A, Nicolas JF: **Capture of a cellular transcriptional unit by a retrovirus: mode of provirus activation in embryonal carcinoma cells.** *J Virol* 1992, **66:**4982-4991.
32. Bonnerot C, Rocancourt D, Briand P, Grimber G, Nicolas JF: **A β-galactosidase hybrid protein targeted to nuclei as a marker for developmental studies.** *Proc Natl Acad Sci USA* 1987, **84:**6795-6799.