

## Review Article

# Regulatory Potential of Mobile Genetic Elements in the Human MGMT Gene

Olga Pidpala and Lukash Lyubov Leonidovna\*

*Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, Kiev, Ukraine*

### Abstract

Mobile Genetic Elements (MGE) make up a large part of the DNA of eukaryotes, in particular, almost 45% of the human genome. Numerous data show the diverse role of these elements in the genome from plasticity factors to mutability or stability. They discuss their role in the evolution of genomes and the evolution of gene regulation. The purpose of this work was to investigate the distribution of MGE in human MGMT gene and their regulatory potential. It has been shown at first that in the human MGMT gene, MGE is present both in the intron sequences and in the promoter region. In the intron sequences, MGEs form composite cluster structures that are the source of various regulatory sequences and have the potential to form alternative promoters. In the promoter region, three sequences of MGE were identified: two LTR repeats and a fragment of the DNA transposon. The MGE fragments in the promoter region of human MGMT also enriched with potential cis-regulatory sequences that may be involved in the regulation of this gene.

**Keywords:** Alternative promoters; Composite cluster structures; Human MGMT gene; Mobile Genetic Elements (MGEs); Promoter; Regulatory elements

\*Corresponding author: Lukash Lyubov Leonidovna, Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, 03680, 150 Zabolotnogo St., Kiev, Ukraine, Tel: +380 445260739; E-mail: lukash.imbg@gmail.com

**Citation:** Pidpala O, Leonidovna LL (2018) Regulatory Potential of Mobile Genetic Elements in the Human MGMT Gene. J Genet Genomic Sci 3: 008.

**Received:** April 15, 2018; **Accepted:** May 24, 2018; **Published:** June 07, 2018

**Copyright:** © 2018 Pidpala O and Leonidovna LL. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Introduction

The eukaryotic genome is a complex and dynamic structure. About 50% of the human genome, and possibly more is covered by Mobile Genetic Elements (MGE), which for a long time were considered unnecessary ballast [1,2]. Studies of recent years convincingly testify to the important role of these elements in the evolution of genomes and in the evolution of gene regulation [3-15]. They are not only mutagenic factors, but they can also be the source of a variety of regulatory sequences, such as sites of alternative splicing, cis-regulatory modules, which are clusters of binding sites of transcription factors, or play the role of alternative promoters [16-24]. MGE tends to integrate into non-coding genome sites (introns, flanked genes and intergenic sites) [25]. On average, in the human introns, MGE is approximately 89.5%; in exons, they account for slightly more than 10%, in particular about 4% for protein-encoding genes [26]. Up to 20% of the genes contain MGE in non-translated regions of mRNA, where they can affect the regulation of gene expressions; in particular, the MGE in 5'UTR affects the initiation of translation [27]. It is known that about 25% of human genes contain MGE in promoter regions, and today there is convincing evidence of their involvement, in particular promoters of retroelements, in the regulation of gene transcriptional activity [21,26]. There are cases where MGE plays the role of alternative promoters, which leads either to increase the level of expression of the corresponding gene, or to change the tissue specificity of its expression [28].

Human MGMT gene encodes a reparative enzyme called O6-methylguanine-DNA methyltransferase, which removes alkylated groups from the O6 position of guanine in DNA and protects cells from their toxic and mutagenic effects. Expression of this gene and the activity of the enzyme themselves have wide limits both in between and intrinsically individual variations, indicating that its regulation is complicated [29]. Given the complexity of this problem, it might be worthwhile to focus on MGE, the role of which in gene regulation has recently been discussed quite widely. Numerous databases today have enormous material on the availability of mobile elements, but the data array is not characterized and not generalized. Perhaps the MGE of MGMT gene is also a source of a variety of regulatory sequences, which led us to focus our attention on the study of this issue.

### Materials and Methods

The nucleotide sequence of the MGMT gene is taken on the Ensembl site. Data on the promoter region of the gene were obtained from GenBank (X61657), about the potential promoter regions of the gene being studied from the AceView database. The results of the search and identification of IHE are done with the CENSOR program. The homology between the sequences studied was determined by the BLAST 2.2.32 program. Functional sites are defined by TFSEARCH: Searching Transcription Factor Binding Sites (ver 1.3). The search for potential regulatory sequences was performed by SITECON, SiteGA, BLASTN 2.2.26 using the TRRD database resources (Transcription Regulatory Regions Database) and CISTER: Cis-element Cluster

Finder, WWW Signal Scan and Tfsitescan. These methods and databases were used in the relevant sections of the article.

## Results and Discussion

### Distribution of MGE in the gene of human MGMT reparative enzyme

Human MGMT gene is localized on telomeric region of chromosome 10 at position 10q26 and consists of one non-coding and four coding exons and four introns. Mobile genetic elements in human MGMT gene are present in all intron sequences and absent in exons (Figure 1). Most of their fragments were detected in the intron 3 (44.74%) and the smallest in the intron 4 (8.26%). In most of the introns, non-LTR retrotransposons predominate among MGE classes (in particular LINE-elements). The share of endogenous retroviruses as well as DNA transposons is insignificant in all introns (Table 1).

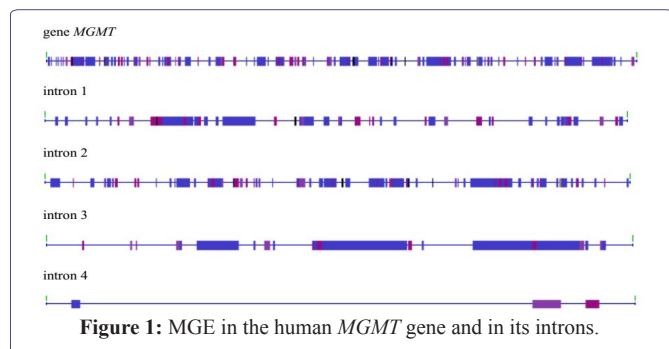


Figure 1: MGE in the human MGMT gene and in its introns.

Intron number (bp)	% MGE, (bp)				
	DNA-transposons	Endogenous retroviruses	LTR retrotransposons	Non-LTR retrotransposons	In total
intron 1 (68.944)	6.05 (4169)	2.36 (1625)	-	19.26 (13281)	27.67 (19075)
intron 2 (171.517)	4.14 (7102)	2.19 (3762)	0.24 (409)	20.75 (35590)	27.32 (46863)
intron 3 (51.158)	2.26 (1158)	2.63 (1345)	-	39.85 (20388)	44.74 (22891)
intron 4 (7.446)	2.25 (167)	4.71 (351)	-	1.3 (97)	8.26 (615)
gene (300824)	4.21 (12676)	2.35 (7083)	0.24 (409)	23.03 (69275)	29.73 (89443)

Table 1: The proportion of MGE in intron sequences of human MGMT gene.

Clusters of MGE were found in the intron sequences of the studied gene. Thus, in intrusion 2 there are three compositional clusters, which, in addition to LINE-elements, include representatives of other classes (Table 2). Intron 3 also revealed three clusters of LINE-elements, one of which is composite (Table 3).

It is known that MGE tends to cluster in intragenic sites of genes and introns [30]. Among the MGEs, clustering tends to Alu-repeats and suggests that these clusters are involved in chromosomal rearrangements [31]. In the case of the MGMT gene of the individual, all

identified clusters include fragments of LINE1-elements alone or in combination with representatives of other classes of MGE (Tables 2 and 3).

Coordinates within the intron	MGE	Length, bp	Class / family	Chain	Number cluster
103739-104018	AluSg	280	NonLTR/SINE/SINE1	-	1
104019-104712	L1ME_ORF2	694	NonLTR/L1	+	1
104713-105077	L1ME_ORF2	365	NonLTR/L1	+	1
125707-125739	MuDRF-2_MLP	33	DNA/MuDR	+	2
125742-125938	L1MC4_Send	211	NonLTR/L1	+	2
125941-126057	MER81	115	DNA/hAT	-	2
126058-126151	L1MC4_Send	101	NonLTR/L1	+	2
150912-151320	MER61E	414	LTR	-	3
151322-152319	L1HS	1005	NonLTR/L1	-	3
152320-152398	L1-2_Cja	79	NonLTR/L1	+	3

Table 2: MGE clusters in human MGMT intron 2.

Coordinates within the intron	MGE	Length bp	Class /family	Chain	Number cluster
13431-14056	L1MEC_5	626	NonLTR/L1	+	1
14060-14677	L1ME_ORF2	618	NonLTR/L1	+	1
14681-14963	AluSz6	283	NonLTR/SINE/SINE1	+	1
24109-24489	L1PA6	378	NonLTR/L1	-	2
24490-30537	L1HS	6049	NonLTR/L1	-	2
30538-30654	L1PA6	120	NonLTR/L1	-	2
30655-31384	L1-2_Cja	733	NonLTR/L1	+	2
42743-42794	L1MC1_EC	52	NonLTR/L1	-	3
42795-45806	L1MB3_EC	3223	NonLTR/L1	-	3
45807-45933	L1MEf_Send	140	NonLTR/L1	-	3
45941-46419	L1MD1_5	476	NonLTR/L1	-	3

Table 3: MGE clusters in human MGMT intron 3.

Recently, the literature actively discusses the role of MGE in regulation of genetic activity. For retrotransposons, in particular for LINE-elements located in intron sequences, it has been shown that they can interfere with the transcription of the gene, causing exonization, introns content, crypt polyadenylation, or act as a “downtrodden policeman” by regulating the passage of RNA polymerase II [32,33]. A “gene breakdown” hypothesis is also proposed, according to which the LINE1-element, which is located in the intron in the opposite direction of transcription, can “break” the transcript into two parts [34].

Consequently, the human MGMT gene of the MGE is present in the intron sequences. In two introns, they form composite cluster structures that can be the source of a variety of regulatory sequences and potentially affect its expression.

### Composite cluster structures of MGE in the introns of the investigated gene as a source of potential regulatory sequences

Of the six clusters we identified, within the introns of the human

MGMT gene, for the further analysis, those that, in addition to the fragments of LINE-elements, also contain Alu-repeats. Such compositional clusters revealed two, one in the intron 2 (Table 2, cluster number 1) and in the 3<sup>rd</sup> (Table 3, cluster number 1). The length of the first cluster is 1339 bp. In addition to the full-length AluSg-repeater, which is in the opposite direction to the transcription, there are fragments of two LINE-elements. Within the AluSg-repeat, homology with binding sites for four transcription factors, namely YY1, p300, C/EBP and for transcription activator of HSF2 heat shock genes has been detected.

Interestingly, the analysis of the consensus sequence of Alu-repeats showed that it contains conservative sites that the YY1 transcription factor can theoretically bind to [35]. It is known that there are large number of consensus sites in the human genome to bind the transcription factor YY1, and some of them (24%) are localized in Alu-repeats. It is noted that about 50% of Alu-repeats contain a potential site for binding YY1 [35]. This allowed expressing the assumption about the role of the Alu-repeats of activation and in the suppression of transcription. The assumption is indirectly confirmed by the data on the ability of the YY1 protein to bind to regulatory regions of the gene and thereby increase or suppress transcription [36]. The YY1 property to attract other proteins to the site of its binding to DNA also can affect the activity of the genome. For example, YY1 specifically binding to the hRPD3 protein, which detects histone deacetylase activity, may initiate complexation in certain regions of the genome of the complex that translates chromatin into an inactive state [37]. Such interaction can lead to changes in the level of transcription. In addition, YY1 protein may interact with ADPRT protein, which has ADP-ribosyl transferase activity [38]. It is shown that YY1 protein, interacting with the protein ADPRT, stimulates its auto-ribosylation [39]. On the basis of these data, the mechanism by which the negative regulation of transcription with the participation of Alu-repeats can be carried out is proposed. The YY1 protein binds to the Alu-repeat “attracts” the ADPRT protein and initiates its auto ribozylation. In a certain region of the gene there is a complex that sterically opposes the formation of the complex of initiation, of transcription. According to the authors in a similar way, the regulation of genes that are under the control of hormone-associated elements can be carried out. This will reveal data on the interaction of the ADPRT protein with the TR/RXR complex, which complicates the binding of nuclear receptors to the complex of initiation [40]. In the consensual sequence of Alu-repeats, a block of AGGTCA with which the transcription factors that belong to the family of nuclear hormonal receptors can be identified. The bulk (more than 70%) of the potential hormone-associated elements for the thyroid hormone, retinoic acid and estrogens is located exactly in the Alu-repeats. CV-1 cells show that binding to nuclear receptors, the Alu-Associated DR-4 element can regulate transcription activity depending on the presence of the thyroid gland hormones [35]. Proteins of a large family of nuclear hormonal receptors can bind not only to the AGGTCA block but also to the various variants that arise due to its duplication. It is suggested that the Alu-repeats the “containers” that contain sets of potential sequences for the binding of various transcription factors [41,42].

Within the fragments of the LINE-elements that are part of the composite cluster structure within the intron 2 MGMT gene, homology with binding sites for 22 transcription factors was detected. Among the number of identified potential sites, I would like to highlight the

sites of Heat Shock protein binding (HSF2), C/EBP, SRY, STAT, Oct, GATA and AP-1. In addition, TATA box and Glucocorticoid Receptor (GR) and ROR alpha 1 (orphan hormone nuclear receptor) binding sites have been identified. This is particularly interesting as previously the presence of sites for glucocorticoid receptor binding within Alu-repeats has been shown [35]. It has also been found that hormone-acceptor elements for the thyroid hormone, estrogen and retinoic acid are mainly localized in Alu-repeats [35].

The presence of TATA box can be a prerequisite for the existence of an alternative promoter in the human MGMT gene within the intron 2. It is worth noting that in addition to the internal promoter, L1 contains an Antisense Promoter (ASP) [43]. The analysis of the database of the expressed gene sequences of the man revealed 49 chimeric transcripts, which begin in LPAA and are part of the mRNA of known genes [44]. In 45 cases, the direction of transcription from the ASP and the promoter of the gene coincided, and the four ASP activities led to the formation of the corresponding complementary RNA. It is assumed that the L1-compatible ASP gene may serve as an alternative promoter and may either lead to the appearance of a chimeric mRNA that is translated to form the same protein (in the case of L1, “topically” relative to the point of transcription), or to the formation of 5'-truncated mRNAs, the translation of which leads to the appearance of various N-terminal forms of protein (in the case of the location of L1 in the intron of the gene). In addition, the ASP activity deserves special attention to be aligned with the L1 gene located in the intron, since in this case, chimeric RNAs that contain sequences complementary to the exons of the gene that potentially can regulate the activity of the corresponding gene through the mechanism of RNA interference are formed.

It is worth mentioning also the “gene-breaking” hypothesis, according to which L1, located in the introns of the gene in the opposite direction of tracing the orientation gene, can “break” the transcript into two parts: 1-the RNA, which covers “above lying” exon and ends in the antisense site polyadenylation L1 and 2-a transcript consisting of an ASP L1 and encompassing “non-existent” exons [45].

The second composite cluster we investigated is located within the intron 3 (Table 3, cluster number 1). It has a length of 1527 bp and consists of two fragments of LINE-elements and a full-length AluSz6-repeat. The transcription direction of all the constituent parts of the cluster composite structure in this case coincides with the direction of transcription of the gene. Interestingly, as in the case of the AluSg-repeat from cluster 1 in intron 2, the AluSz6-repeat also contains an element of response to the HSF2 heat shock proteins. Within the fragments of LINE-elements among other potential sites, as in the previous cluster, the homology with binding sites for C/EBP, SRY, STAT, Oct, GATA and AP-1, TATA box and sites for binding to ROR alpha 1 (orphan hormone nuclear receptor).

As can be seen from the results presented in table 4, for some transcription factors, sequences that are homologous to their binding sites are present in retroelements belonging to different families. There are those that are present only in the sequence of fragments of LINE-elements, including the TATA box and the potential hormone-associated element for the retinoid orphan receptor ROR alpha 1 (orphan hormone nuclear receptor).

Transcription factor, name	Ingredients of composite clusters					
	intron 2 (cluster 1)			intron 3 (cluster 1)		
	AluSg	L1ME_ORF2	L1ME_ORF2	L1MEC_5	L1ME_ORF2	AluSz6
C/EBP	+	+	+	+	+	
1-Oct		+	+	+	+	+
SRY		+	+	+	+	+
GATA		+	+	+	+	+
HSF2	+	+	+			+
AP-1		+		+		+
Pbx-1		+		+	+	
MZF1		+	+	+		
TATA		+			+	
STAT			+		+	
RORalp		+			+	

**Table 4:** Representation of potential sites of binding the boundaries of composite cluster structures of MGE in human MGMT gene introns.

Thus, having analyzed the two compositional cluster structures in the intron 2 and intron 3 the human MGMT gene, which include Alu-repeats and fragments of LINE-elements, it was found that both Alu-repeats contain sequences that are homologous to the response elements to the heat shock proteins, and cluster structures within the fragments of LINE-elements have TATA box sequences. This gives grounds for considering analyzed composite cluster structures of MGE within the intron sequences of human MGMT gene as potential alternative promoters.

### Motives of regulatory sequences in the promoter regions of the MGMT gene within the MGE

The promoted gene promoter (X61657) has a length of 1157 bp and covers exon 1 part intron 1. It is devoid of TATA or CAAT sequences, contains CG-rich regions and is structurally reminiscent of genes of the household. It also contains SP1, AP-1 and AP-2, NF-kapB sites, two elements that bind the Glucocorticoid Receptor (GRE) and a 59-bp size element, which is located on the first eccentric-intron boundary, necessary for effective transcription of the reporter designs [46,47].

### Fragments of MGE in the promoter region of the investigated gene

By analyzing the promoter region of the human MGMT gene, sequences of two fragments of LTR repeats of mammalian retroviruses were identified, namely, LTR-repeats of endogenous retroviruses ERV3 MLT1C2 and MLT1C, which are located in the distal part of the promoter sequence (Table 5) and are about 23% of the total length of the promoter. It should be noted that the described response elements for Glucocorticoid hormones (GRE) with coordinates 28-42 and 63-77 are localized within one of the fragments of the LTR-repeat, namely, MLT1C2 [46].

Referential promoter, symbolic	Data on MGE				
	Element	Class	Length	Direction	Coordinates within promoter
	MLT1C2	ERV/ERV3	117	c	-949/-839
hO <sup>o</sup> P	MLT1C	ERV/ERV3	147	c	-813/-667
	SETARIA1	DNA	341	D	-265/+76

**Table 5:** Fragments of MGE in the promoter region of human MGMT gene.  
**Notes:** c - complementary; d - straight.

In addition, a sequence of the fragment of the Mutator-like non-autonomous DNA transposon of SETARIA1 was found (Table 5). Interestingly, the minimum promoter (886-955) and the sequence of SP1 of the site (862-867) are within the given sequence SETARIA1 [46]. Thus, in the case of the human MGMT gene, the promoter region contains fragments of three MGE, which is almost half its length.

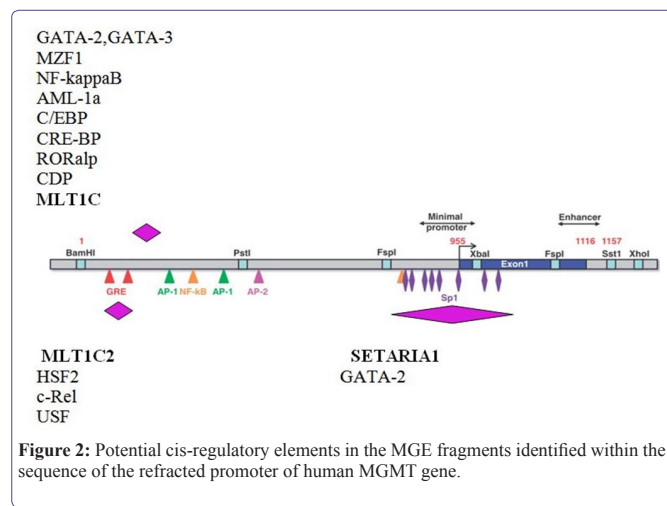
It is known that up to 25% of genes contain MGE in promoter regions [21]. Especially enriched MGE genes that are associated with metabolic processes [48]. For human DNA se II and CAML genes, it has been shown that the presence of Alu-repeats in the promoter may affect the expression of these genes [49]. It is known that the human gene MSLN (Mesothelin, a megakaryocyte-potentiating factor) has two promoters, one of which is formed by the sequence of LTR, and the other is an MIR element (tRNA-like SINE) [27]. The only currently known promoter of the BAAT gene specifically expressed in the liver and involved in the development of a hereditary disease associated with bile metabolism, is also the sequence of LTR-repeats [50]. An interesting case of regulation of the transcription of the NAIP gene encoding one of the inhibitors of apoptosis is interesting. It has been shown that the promoter portions of this human gene and rodents do not have homology, but in one and the other, LTR are alternative promoters [51]. In humans, the integration of LTR has led to the formation of a tissue-specific promoter that is active mainly in the testicles, whereas in rodents two LTRs are described that are capable of initiating transcription. One of them is the main, constitutive promoter, active in all rodents, and the other an alternative, found only in the mouse. It is important to note that LTRs of humans and rodents that are in the promoter region of the NAIP gene are not related. This case is an example of the independent involvement of LTR in regulating the transcription of orthologic genes [51].

### Potential cis-elements within the fragments of mobile genetic elements in the promoter regions of the investigated gene

Among the potential cis-regulatory sequences within the LTR-repeats, an element of response to the HSF2 heat shock proteins, binding sites for several representatives of the GATA family of proteins (important regulators specification and differentiation of various tissues), MZF1 (involved in the control of cell proliferation and carcinogenesis), NF-kappaB (participation in the activation of transcription of numerous cytokine and immunoregulatory genes), AML-1 (regulation of hemopoiesis, angiogenesis and neurogenesis), C/EBP (control



of the differentiation of various cell types and a key role in regulating cellular proliferation through interaction with cell cycle proteins), CRE-BP (an important role in the development and functioning of the nervous system), Nkx-2.5 (the regulation of the expression of tissue-specific genes, the development of the heart, the time and spatial development models), CDP (regulates gene expression, morphogenesis and differentiation, the role in the progression of the cell cycle, especially in the S-phase) and the potential hormone-associated element for the retinoid orphan R receptor ROR alpha 1 (orphan hormone nuclear receptor) (Figure 2).



As part of the sequence of the DNA fragment of the transposition of SETARIA1 (Figure 2), apart from the numerous potential binding sites for the Sp1 factor (one of the major transcription factors involved in the regulation of the cell cycle, chromatin structure changes and DNA methylation regulation), the motives for recognition for transcription factors GATA-2 (hematopoietic transcription factor), c-Rel (participation in immune and inflammatory reactions, developmental processes, cell growth and apoptosis) and USF (one of the key regulatory elements of gene expression).

Recently, not only for LTR elements, called “regulatory information packages”, but also for other MGEs, the presence of numerous regulatory sequences in their structure has been shown [52]. Integrating in the promoter regions of the genes, MGEs can be the source of cis-regulatory modules, which are clusters of transcription factor binding sites [21,48,53]. In particular, in the consensus Alu-repetition sequence, there were found binding sites for 20 transcription factors, the functional activity of most of which has been proved experimentally [48]. In addition, Alu-repeats identified functional binding sites for retinoic acid receptors and hormone-acceptor elements [54-57]. As already noted, the bulk of the potential hormone-associated elements for the thyroid hormone, retinoic acid and estrogens are located exactly in the Alu-repeats [35].

The direct involvement of Alu-repeats in expression regulation is shown for genes that are associated with differentiation and development, namely for PTH, FcεRI-γ, CD8α, CHRNA3, BRCA-1 and PLOD-1 genes [58]. In Alu-repeats, sites of transcription factor binding have been identified which are involved in hemopoiesis, T-cell-differentiation, and the development of various organs (eyes, teeth, heart, lungs and brain), which is another evidence of participation Alu-repeats in ontogenesis [48].

Thus, in the referenced promoter of the MGMT gene of the human, three sequences have been identified that are homologous to the MGE fragments, namely two LTR-repeats and a fragment of the DNA-transposon. The fact that one of the LTR-repeats contains the previously described response elements for Glucocorticoid hormones (GRE) and that the known minimal promoter and sequence of the Sp1 site are located within the fragment of the DNA-transposon, confirms the important role of MGE in gene regulation. Also, the identified MGE fragments in the promoter region of human MGMT enriched with potential cis-regulatory sequences that may be involved in the regulation of this gene.

### Fragments of the MGE as components of potential alternative promoters

In addition to the referenced promoter, the AceView database contains information on eight potential promoter sites (length of 2,000 bp) for the gene in question. For four, there is information that these sequences may contain a promoter (aAug10, cAug10, eAug10 and iAug10). According to BLAST, the analysis of potential promoter regions of the human MGMT gene revealed that of the eight potential promoter regions, three sequences (cAug10, hAug10 and iAug10) had reflex sequence homology and no five homology sequences. In particular, the dAug10 sequence is localized within the intron 1. Four other sequences (aAug10, eAug10, fAug10 and gAug10) are located within the intron 2.

### Fragments of the MGE in alternative promoter regions of the investigated gene

It has been shown that four of the eight potential promoter regions studied contain fragments of MGE (Table 6 and Figure 3). It is interesting that these are the exact sequences that may include the promoter, as indicated in the AceView database. Therefore, these sequences were analyzed in detail. Two sequences of possible alternative promoters (cAug10 and iAug10) overlapping with the referenced sequence (hO6P) contain additional fragments of MGE-MIRc (Non-LTR-retrotransposon) and MER117 (DNA-transposon), whereas the sequences of two fragments of LTR-repeats of mammalian retroviruses, namely the LTR-repeat of the MaLR retrovirus-like element (MLT1C) and the LTR-repeat of the endogenous retrovirus ERV3 (MLT1C2) identified by the CENSOR program as one fragment of the endogenous retrovirus due to the larger size of the nucleotide sequence. This feature of the program should be taken into account in further research. In the sequences aAug10 and fAug10, located within the intron 2, fragments of MGE of different families were identified (Figure 3). Particular attention deserves two sequences: Tigger15a (DNA-transposon) and AluSx1 (NonLTR-Retrotransposon), which is specific to MGE for mammals and primates.

### Potential regulatory sequences within identified MGE fragments

All of the MGE fragments we investigated in the alternative human MGMT gene promoters are enriched in a variety of regulatory sequences (Table 7). In particular, the AluSx1-repeat contains a number of new, besides the described, transcription factors binding sites, as well as tissue-specific promoter, enhancer and seylener sequences (Table 7) [48,54-57]. It is known that Alu-repeats can affect cellular processes by providing new transcription termination sites or splice sites or acting as alternative promoters [16,19,21,53]. Such situations

may interfere with the normal functioning of the gene, but may sometimes lead to the formation of proteins with new functions [59]. DNA transposons are also enriched by regulatory sequences [60]. The fragment of the Tigger15a element we examined contains binding sites for a number of transcription factors (ERE, HRE, MEF-2 and C/EBP), as well as promoter, enhancer and seylelange sequences and the locus control region (locus control region-like region).

Symbolic designation for the promoter	Data on MGE				
	Element	Class	Length	Direction	Coordinates within promoter
aAug10	Tx1-11_Crp	NonLTR/Tx1	60	c	320-379
	Tigger15a	DNA/Mariner	105	c	448-552
	Gypsy-54_PIT-1	LTR/Gypsy	54	d	1278-1331
	DNA4-8_CGi	DNA	104	d	1425-1528
cAug10	MIRc	NonLTR/SINE/SINE2	53	c	115-167
	MER117	DNA/hAT	120	d	202-321
	MLT1C2	ERV/ERV3	397	c	771-1167
	SETARIA1	DNA	341	d	1576-1916
fAug10	ERV-1_Crp-I	ERV	38	c	19-56
	AluSx1	NonLTR/SINE/SINE1	285	c	374-658
iAug10	Gypsy-57_GR-I	LTR/Gypsy	44	c	1777-1820
	MIRc	NonLTR/SINE/SINE2	53	c	208-260
	MER117	DNA/hAT	120	d	295-414
	MLT1C2	ERV/ERV3	397	c	864-1260
	SETARIA1	DNA	314	d	1685-1998

Table 6: Sequence of MGE in potential alternative promoter regions of the MGMT human reparative enzyme gene.

Notes: c - complementary; d - straight.

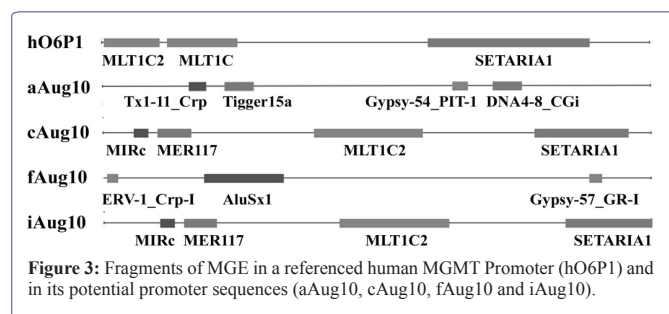


Figure 3: Fragments of MGE in a referenced human MGMT Promoter (hO6P1) and in its potential promoter sequences (aAug10, cAug10, fAug10 and iAug10).

The human genome is not only specific to mammals, primates or humans, or their fragments or copies, but also fragments of MGE of other organisms: animals, plants and even bacteria [61-64]. Such sequences are also enriched in a variety of regulatory elements and can affect the transcriptional activity of the human MGMT reparative gene. It is worth noting that the presence of TATA box within the

MGE of the potential promoter region of fAug10 (indicated in bold in Table 7) may be a prerequisite for the existence of an alternative promoter.

Conditional designation of the promoter	Mobile genetic elements		Regulatory sequences	
	Name	Species affiliation	Linking sites for transcription factors	Regulatory elements
aAug10	Tigger15a	Mammalia	ERE; HRE; MEF-2; C/EBP	promoter; enhancer; silencer; Locus Control Region-like region
fAug10	AluSx1	Primates	<b>TATA box</b> ; YY1; GATA; SF1; Sp1; CAAT box; OCT1; SREBP; PAX2; NF-kappaB; PBX2; SOX9; SRY; WT1; AP-2; C/EBP; P53; CRE; CREB; MEF-2; nCaR-E; TRalpha; RARalpha/RXR; IFN-stimulated response element	promoter; myeloid-specific promoter; erythroid-specific promoter; enhancer; lymphocyte-specific enhancer; H <sub>2</sub> O <sub>2</sub> -inducible enhancer; non-renal silencer cytokine responsive element; Ets-responsive region PMA and cAMP responsive region

Table 7: Regulatory potential sequences of MGE, inherent in the genome of the man, within the potential promoters of the human MGMT gene.

Thus, analyzing the nucleotide sequences of the potential alternative promoter regions of the human MGMT gene from the AceView database revealed that the two sequences (cAug10 and iAug10) are overlapped with the reflux promoter and the other two sequences (aAug10 and fAug10) are located within the intron 2. All of them contain in its composition fragments of MGE that are enriched in a variety of regulatory sequences and can affect the regulation of the transcriptional activity of the human MGMT gene.

There is no doubt that MGE can affect the expression of eukaryotic genes. First and foremost, these processes involve their promoters and their associated regulatory sequences. A global analysis of retrotransposon expression in the human genome revealed ~ 275,000 TSS, which is ~ 31% of all known TSS human genomes, although their level of activity is significantly lower than that of conventional genes. Transcription of sequences of MGE also affects the transcript of encoding proteins [65]. It is shown that 576 promoters of human retrotransposons or their fragments are used as alternatives to the transcription of known genes. Also described are cases of enhancers and MGE infectious agents in the transcriptional networks of human genomes, animal plants [66]. All these facts indicate the important role of MGE in phylogeny and ontogenesis of eukaryotes, but their global significance still needs confirmation [67].

### Conclusion

In the human MGMT gene, MGE is present both in the intron sequences and in the promoter region. In the intron sequences, MGEs form composite cluster structures that are the source of various regulatory sequences and have the potential to form alternative promoters. In the promoter region, three sequences of MGE were identified: two LTR-repeats and a fragment of the DNA-transposon. The fact that

one of the LTR-repeats contains the described response elements of the Glucocorticoid hormones (GRE) and that the minimal promoter and sequence of the SP1 site are located within the fragment of DNA-transposon confirms the important role of MGE in gene regulation. In addition, the MGE fragments in the promoter region of human MGMT enriched with potential cis regulatory sequences that may also be involved in the regulation of this gene.

## References

- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.
- de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD (2011) Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* 7: 1002384.
- Deininger PL, Moran JV, Batzer MA, Kazazian HH Jr (2003) Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev* 13: 651-658.
- Hedges DJ, Batzer MA (2005) From the margins of the genome: mobile elements shape primate evolution. *Bioessays* 27: 785-794.
- Werren JH (2011) Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci USA* 108: 10863-10870.
- Piskurek O, Jackson DJ (2012) Transposable elements: From DNA parasites to architects of metazoan evolution. *Genes (Basel)* 3: 409-422.
- Chalopin D, Naville M, Plard F, Galiana D, Volff JN (2015) Comparative analysis of Transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol Evol* 7: 567-580.
- Platt RN, Vandeweghe MW, Ray DA (2018) Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res* 26: 25-43.
- Medstrand P, vande Lagemaat LN, Dunn CA, Landry JR, Svenback D, et al. (2005) Impact of transposable elements on the evolution of mammalian gene regulation. *Cytogenet Genome Res* 110: 342-352.
- Feschotte C (2008) Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 9: 397-405.
- Jurka J (2008) Conserved eukaryotic transposable elements and the evolution of gene regulation. *Cell Mol Life Sci* 65: 201-204.
- Bourque G (2009) Transposable elements in gene regulation and in the evolution of vertebrate genomes. *Curr Opin Genet Dev* 19: 607-612.
- Rayan NA, Del Rosario RCH, Prabhakar S (2016) Massive contribution of transposable elements to mammalian regulatory sequences. *Semin Cell Dev Biol* 57: 51-56.
- Chuong EB, Elde NC, Feschotte C (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* 18: 71-86.
- Sundaram V, Wang T (2018) Transposable Element Mediated Innovation in Gene Regulatory Landscapes of Cells: Re-Visiting the "Gene-Battery" Model. *Bioessays* 40.
- Callinan PA, Batzer MA (2006) Retrotransposable elements and human disease. *Genome Dyn* 1: 104-115.
- Chenais B (2015) Transposable elements in cancer and other human diseases. *Curr Cancer Drug Targets* 15: 227-242.
- Kazazian HH Jr, Moran JV (2017) Mobile DNA in Health and Disease. *N Engl J Med* 377: 361-370.
- Sorek R, Ast G, Graur D (2002) Alu-Containing Exons are Alternatively Spliced. *Genome Res* 12: 1060-1067.
- Stower H (2013) Alternative splicing: Regulating Alu element 'exonization'. *Nat Rev Genet* 14: 152-153.
- Jordan IK, Rogozin IB, Glazko GV, Koonin EV (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet* 19: 68-72.
- Thornburg BG, Gotea V, Makalowski W (2006) Transposable elements as a significant source of transcription regulating signals. *Gene* 365: 104-110.
- Rebollo R, Romanish MT, Mager DL (2012) Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu Rev Genet* 46: 21-42.
- Batut P, Dobin A, Plessy C, Carninci P, Gingeras TR (2013) High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome Res* 23: 169-180.
- Prak ET, Kazazian HH Jr (2000) Mobile elements and the human genome. *Nat Rev Genet* 1: 134-144.
- Nekrutenko A, Li WH (2001) Transposable elements are found in a large number of human protein-coding genes. *Trends Genet* 17: 619-621.
- van de Lagemaat LN, Landry JR, Mager DL, Medstrand P (2003) Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet* 19: 530-536.
- Faulkner GJ, Carninci P (2009) Altruistic functions for selfish DNA. *Cell Cycle* 8: 2895-2900.
- Christmann M, Verbeek B, Roos WP, Kaina B (2011) O(6)-Methylguanine-DNA methyltransferase (MGMT) in normal tissues and tumors: enzyme activity, promoter methylation and immunohistochemistry. *Biochim Biophys Acta* 1816: 179-190.
- Jurka J, Kohany O, Pavlicek A, Kapitonov VV, Jurka MV (2004) Duplication, coclustering, and selection of human Alu retrotransposons. *Proc Natl Acad Sci USA* 101: 1268-1272.
- Kolomietz E, Meyn MS, Pandita A, Squire JA (2002) The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. *Genes Chromosomes Cancer* 35: 97-112.
- Kaer K, Branovets J, Hallikma A, Nigumann P, Speek M (2011) Intronic L1 retrotransposons and nested genes cause transcriptional interference by inducing intron retention, exonization and cryptic polyadenylation. *PLoS One* 6: 26099.
- Kaer K, Speek M (2012) Intronic retroelements: Not just "speed bumps" for RNA polymerase II. *Mob Genet Elements* 2: 154-157.
- Wheelan SJ, Aizawa Y, Han JS, Boeke JD (2005) Gene-breaking: A new paradigm for human retrotransposon-mediated gene evolution. *Genome Res* 15: 1073-1078.
- Oei SL, Babich VS, Kazakov VI, Usmanova NM, Kropotov AV, et al. (2004) Clusters of regulatory signals for RNA polymerase II transcription associated with Alu family repeats and CpG islands in human promoters. *Genomics* 83: 873-882.
- Thomas MJ, Seto E (1999) Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? *Gene* 236: 197-208.
- Yang WM, Inouye C, Zeng Y, Bearss D, Seto E (1996) Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. *Proc Natl Acad Sci USA* 93: 12845-12850.

38. Oei SL, Griesenbeck J, Schweiger M, Babich V, Kropotov A, et al. (1997) Interaction of the transcription factor YY1 with human poly(ADP-ribosyl) transferase. *Biochem Biophys Res Commun* 240: 108-111.
39. Griesenbeck J, Ziegler M, Tomilin N, Schweiger M, Oei SL (1999) Stimulation of the catalytic activity of poly(ADP-ribosyl) transferase by transcription factor Yin Yang 1. *FEBS Lett* 443: 20-24.
40. Kakizawa T, Miyamoto T, Ichikawa K, Kaneko A, Suzuki S, et al. (1999) Functional interaction between Oct-1 and retinoid X receptor. *J Biol Chem* 274: 19103-19108.
41. Umesonu K, Murakami KK, Thompson CC, Evans RM (1991) Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors. *Cell* 65: 1255-1266.
42. Mangelsdorf DJ, Evans RM (1995) The RXR heterodimers and orphan receptors. *Cell* 83: 841-850.
43. Speek M (2001) Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol* 21: 1973-1985.
44. Matlik K, Redik K, Speek M (2006) L1 antisense promoter drives tissue-specific transcription of human genes. *J Biomed Biotechnol* 2006: 71753.
45. Wheelan SJ, Aizawa Y, Han JS, Boeke JD (2005) Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Res* 15: 1073-1078.
46. Harris LC, Potter PM, Tano K, Shiota S, Mitra S, et al. (1991) Characterization of the promoter region of the human O6-methylguanine-DNA methyltransferase gene. *Nucleic Acids Res* 19: 6163-6167.
47. Harris LC, Remack JS, Brent TP (1994) Identification of a 59 bp enhancer located at the first exon/intron boundary of the human O6-methylguanine DNA methyltransferase gene. *Nucleic Acids Res* 22: 4614-4619.
48. Polak P, Domany E (2006) Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* 7: 133.
49. Usmanova NM, Kazakov VI, Tomilin NV (2008) [Retroposons of Alu family from cis-regulatory modules of DNase II and CAML genes effect gene expression in A549 and HEK293 cells]. *Tsitologija* 50: 249-255.
50. Bannert N, Kurth R (2004) Retroelements and the human genome: new perspectives on an old relation. *Proc Natl Acad Sci USA* 101: 14572-14579.
51. Romanish MT, Lock WM, van de Lagemaat LN, Dunn CA, Mager DL (2007) Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. *PLoS Genet* 3: 10.
52. Sverdlov ED (1998) Perpetually mobile footprints of ancient infections in human genome. *FEBS Lett* 428: 1-6.
53. Shankar R, Grover D, Brahmachari SK, Mukerji M (2004) Evolution and distribution of RNA polymerase II regulatory sites from RNA polymerase III dependant mobile Alu elements. *BMC Evol Biol* 4: 37.
54. Vansant G, Reynolds WF (1995) The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. *Proc Natl Acad Sci USA* 92: 8229-8233.
55. Laperriere D, Wang TT, White JH, Mader S (2007) Widespread Alu repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution. *BMC Genomics* 8: 23.
56. Babich V, Aksenov N, Alexeenko V, Oei SL, Buchlow G, et al. (1999) Association of some potential hormone response elements in human genes with the Alu family repeats. *Gene* 239: 341-349.
57. Piedrafita FJ, Molander RB, Vansant G, Orlova EA, Pfahl M, et al. (1996) An Alu element in the myeloperoxidase promoter contains a composite SP1-thyroid hormone-retinoic acid response element. *J Biol Chem* 271: 14412-14420.
58. Hamdi HK, Nishio H, Tavis J, Zielinski R, Dugaiczuk A (2000) Alu-mediated phylogenetic novelties in gene regulation and development. *J Mol Biol* 299: 931-939.
59. Romanish MT, Nakamura H, Lai CB, Wang Y, Mager DL (2009) A Novel Protein Isoform of the Multicopy Human NAIP Gene Derives from Intra-genic Alu SINE Promoters. *PLoSOne* 4: 5761.
60. Bolotin E, Chellappa K, Hwang-Verslues W, Schnabl JM, Yang C, et al. (2011) Nuclear receptor HNF4 $\alpha$  binding sequences are widespread in Alu repeats. *BMC Genomics* 15: 560.
61. Pidpala OV, Lukash LL (2015) Motives of regulatory sequences in human promoter regions of human MGMT within the framework of mobile genetic elements. *Factors of experimental evolution of organisms* 16.
62. Mayorov VI, Rogozin IB, Elisaphenko EA, Adkison LR (2000) B2 elements present in the human genome. *Mamm Genome* 11: 177-179.
63. Pidpala OV, Yatsishina AP, Lukash LL (2006) Mobile genetic elements in the genome-oncosuppressor p53: evolutionary aspect. *Factors In Experimental Evolution of Organisms* 3: 43-48.
64. Pidpala OV, Yatsishina AP, Lukash LL (2007) Fragments of bacterial IS-elements and mobile genetic elements of eukaryotes in human mt DNA. *Achievements and problems genetics, breeding and biotechnology* 1: 498-502.
65. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, et al. (2009) The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* 41: 563-571.
66. de Souza FS, Franchini LF, Rubinstein M (2013) Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Mol Biol Evol* 30: 1239-1251.
67. Patrushev LI, Kovalenko TF (2014) Functions of noncoding sequences in mammalian genomes. *Biochemistry (Mosc)* 79: 1442-1469.