UNIVERSITÀ DI PARMA

UNIVERSITY OF PARMA

PhD in Biotechnology and Life Science

CICLO XXXII

Insights into *Alu* retrotransposons: mechanism of *Alu* transcriptome alteration in response to virus infection and novel effects on gene expression

Coordinator:

Prof. Simone Ottonello

Tutor:

Giorgio Dieci

PhD student: Simona Cantarella

Years 2016/2019

Abstract

A large portion of the human genome is composed of repeated sequences, with *Alu* retrotransposons representing the most abundant repetitive elements. *Alu* sequences belong to the class of the Short Interspersed Nuclear Elements (SINEs) and depend on the Long Interspersed Nuclear Elements (LINEs) for their mobilization into the genome. The efficiency of *Alu* amplification during primate evolution suggests a positive driving force for their accumulation, bringing up to 1 million copies in the human genome. For unclear reasons, the majority of *Alu* sequences is repressed by tight epigenetic silencing, which is released in response to cell stresses such as virus infection and cancer progression. Adenovirus 5 (Ad5) is known to cause an increase of *Alu* transcription in HeLa, myelogenous leukemia and embryonic kidney cell lines, even though the virus factors that are responsible for this transcriptional enhancement have not been identified yet. Potential candidates could be represented by oncovirus proteins that induce a global remodeling of the host epigenetic landscape. For example, the Adenovirus early E1A protein interacts with the host tumor suppressor Rb, the lysine acetylase p300 and the p400 ATP-dependent chromatin remodeling complex, resulting in the induction of quiescent fibroblasts to enter the S-phase of the cell cycle.

The exceptional success of *Alu* expansion and their retention even at the cost of a strong epigenetic silencing, which is released by virus infection, led us to investigate the molecular mechanism of *Alus* activation and their potential involvement in various cell processes.

Firstly, genome-wide *Alu* profiling was performed in quiescent human primary fibroblasts infected with the Ad5 *dl*1500 mutant, which only expresses the oncovirus small E1A protein. A total of 1880 Polymerase III (Pol III) transcribed *Alu*s were detected through high throughput RNA-sequencing (RNA-seq), revealing a 4-fold increase of the average *Alu* expression induced by small E1A.

With the aim of identifying small E1A-host protein interactions that are crucial for the activation of *Alu* expression, the host proteins Rb, p300 and p400 were put under investigation. *Alu* expression profiling was performed in cells infected with small E1A mutants that are not capable to bind Rb, p300 or p400. RNA-seq and RT-qPCR data revealed that the small E1A-p400 binding mutant was the least efficient in activating *Alu* expression, whereas a milder effect was reported for small E1A-Rb and small E1A-p300 binding mutants. Moreover, ChIP-sequencing (ChIP-seq) analyses of *dl*1500 infected fibroblasts revealed an enrichment of H3K4me1 within the body of small E1A-induced *Alu*s. Our data point toward the existence of H3K4me1 *Alu* loci that are recognized by p400 bound by small E1A, resulting in *Alu* transcription in response to virus infection.

Secondly, two *Alu* sequences were stably overexpressed in one primary and one cancer cell line (human fibroblasts and HeLa cells, respectively) and differential gene expression in *Alu*-overexpressing cells was performed through RNA-seq data analysis. Among the two *Alu*- overexpressing fibroblast cell lines, 330 genes were detected as Differentially Expressed (DE), whereas only two genes were differentially expressed in HeLa cells. Interestingly, DE genes detected in *Alu*-overexpressing fibroblasts were significantly enriched in pathways belonging to cell cycle progression and mitotic entry. The promotion of cell cycle was also supported by a significantly higher percentage of cells in the S-phase compared to control samples as revealed by flow cytometry.

The studies conducted in this work identify Ad5 small E1A as one of the virus factors that enhance *Alu* transcription, possibly through binding with the p400 complex. Interestingly, the overexpression of two *Alu* sequences in human fibroblasts leads to

the stimulation of cell cycle progression, resulting in the same phenotype as previously observed in Ad5 infected fibroblasts. This brings us to speculate that the overexpression of *Alu* sequences during stress response to viral infection could be exploited by Ad5 to sustain cell proliferation.

Acknowledgements

First, I would like to thank my advisor Prof. Giorgio Dieci for welcoming me to work in his laboratory and for his help in gaining experience and knowledge on the exciting field of non-coding RNA research. Moreover, I will always be thankful for the great opportunity, for the fundamental support and for the valuable confidence that Prof. G. Dieci gave me in conducting part of this research at the University of Los Angeles, California (UCLA).

A very special thanks goes to Prof. Barbara Montanini for her precious mentoring and constant support. With her dynamicity and insightful comments, she has represented a stable point of reference and a motivational source over these last three years.

I also thank Davide Carnevali for performing RNA-sequencing data analysis.

Thanks to Elisabetta Levati for her patience and constant advices.

Many thanks are due to Prof. Matteo Pellegrini for welcoming me in his research group at the Department of Molecular Cell and Developmental Biology (UCLA), allowing me to acquire skills in bioinformatic data analysis; and many thanks to Prof. Arnold J. Berk for hosting me in his group at the Molecular Biology Institute (UCLA), giving me fundamental knowledge in gene regulation.

I would like to particularly thank Marco Morselli for his extreme kindness in guiding me through my whole experience in Los Angeles, since the first days of my arrival to the very end. Especially, I must thank him for his very patient and professional supervision during my stay at UCLA, and for teaching me and performing part of the RNA-sequencing libraries necessary to conduct this research.

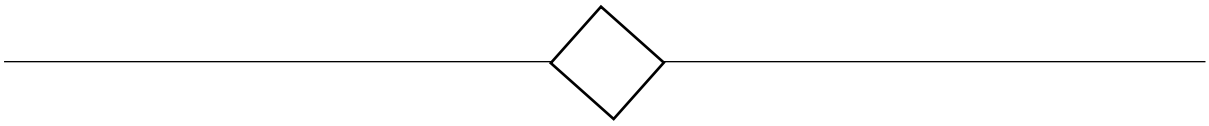
I thank all the members of the Pellegrini and Berk groups: thanks to Anela Tosevska for being such a supportive friend and colleague, sharing countless and always fun experiences inside and outside the laboratory; thanks to Giorgia Del Vecchio for her

endless patience in giving me advices on cell culture and data analysis, and for her continuous encouragement. Thanks to Mila Rubbi for her psychological and professional support, helping me to overcome the numerous difficulties I encountered during my experience abroad. I also thank Nathan Zemke for his always inspiring comments and expertise in virus infections and cell culture.

A very big thanks goes to my fellows and colleagues from the University of Parma: Stefano, Sara, Ilaria, Valentina, Anastasia, Alessia and Davide. Thank you for being such an amazing group of friends, making my time in the lab always fun and enjoyable, even in the hardest times.

Last but not least, I am especially grateful to my family. There are no words to describe how thankful I am for giving me all the support and unconditional trust throughout my entire path of studies. Your support will be always fundamental.

Table of contents



Chapter I

Introduction

Retrotransposons of the human genome

The sequencing of the human genome, which was concluded in 2003 in the context of the Human Genome Project, has allowed to obtain valuable information on its structure and evolution. From analyses of the initial draft sequence, it is evident that only a small portion (1.5%) is comprised of protein-coding genes, whereas transposable elements account for about half of the genome (Figure 1) [1].

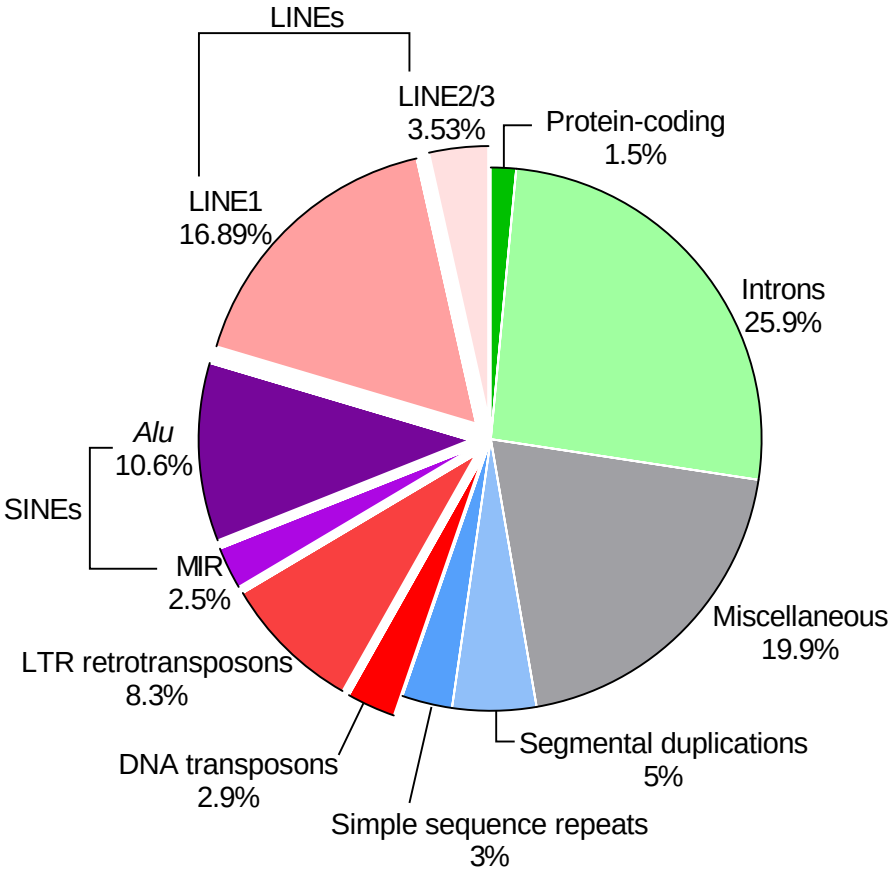


Figure 1 Composition of the human genome. Roughly 45% of the human genome is composed of transposable elements, which are represented as exploding wedges. Data are derived from [1].

A clear function of transposable sequences has not been identified yet, although their retention throughout human evolution has probably a functional significance. Indeed, mobile sequences represent an important source of genetic variation through various

mechanisms such as genome structural rearrangements [2], the creation of new genes [3] and the acquisition of regulatory functions that influence gene expression [4].

Two alternative processes are responsible for transposons expansion: the "copy-and-paste" mechanism consists in the synthesis of an RNA intermediate that is subsequently retrotranscribed and reinserted into a different genomic locus (RNA transposons, or retrotransposons); the "cut-and-paste" mechanism involves the direct excision and re-insertion of the transposable element without the synthesis of an intermediate RNA species (DNA transposons) (Figure 2A and 2B).

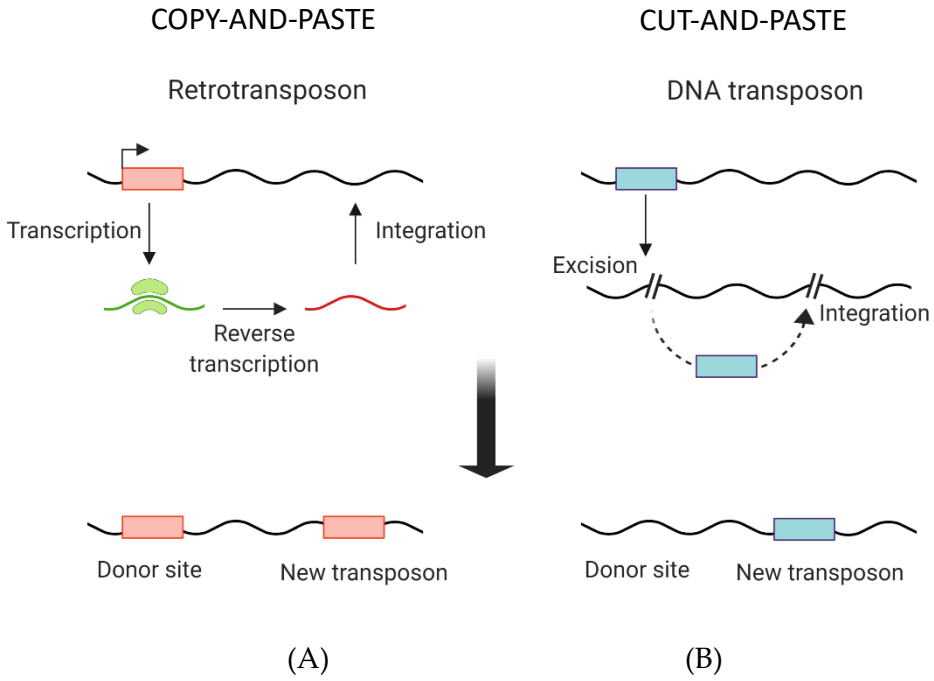


Figure 2. Different mechanisms of transposon amplification into the genome. The copy-and-paste process takes place during retrotransposon amplification through the creation of an RNA intermediate (A), whereas DNA transposons are re-inserted through a more straightforward cut-and-paste process (B).

Retrotransposons are the most abundant mobile elements of the human genome, being present at almost 3 million copies. These sequences fall into one of three types: Retrovirus-like elements (LTR retrotransposons), Long Interspersed Nuclear Elements (LINEs) and Short Interspersed Nuclear Elements (SINEs) (Figure 3).

Long Terminal Repeats (LTR) are flanked by 100-300 bp direct terminal repeats and bear the *gag* and *pol* genes, which are necessary for the retrotransposition process. The similarity of LTRs to retroviruses is remarkable. Indeed, it is though that retroviruses have arisen from LTR elements by acquisition of the gene coding for the protein forming the viral envelope (*env*) [5,6]. Most of the endogenous retroviruses inserted into the human genome more than 25 Million years ago (Mya), and it is thought that LTR have lost retrotranspositional activity in the human genome [1,7,8].

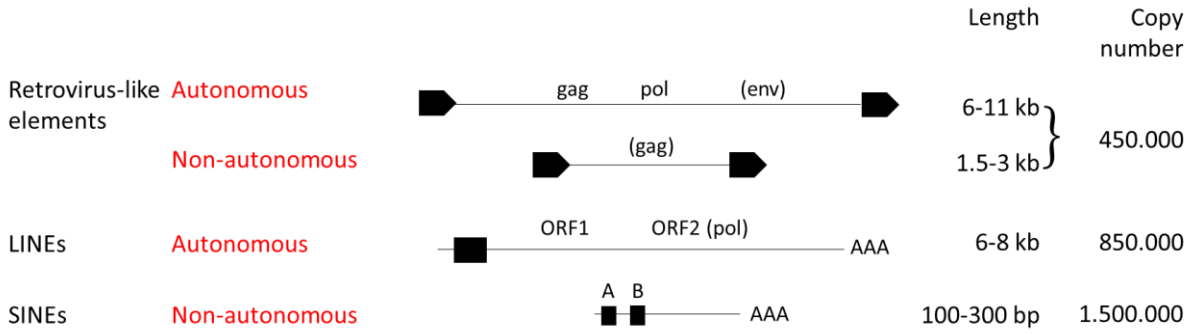


Figure 3. Retrotransposons of the human genome. In mammals, most of the LTR elements have lost the internal sequence by homologous recombination between the flanking LTR sequences. Adapted from [1].

LINE1 and *Alu* elements are the only active transposons of the human genome.

Based on their evolutionary history, human LINEs are divided in three classes (LINE1, LINE2 and LINE3), with LINE1 being the most abundant elements (17% of the human genome). LINEs are about 6-kb in length and harbour an internal Polymerase II (Pol II) promoter, which drives the transcription of the protein coding genes ORF1 and ORF2. These proteins have reverse transcription, chaperone and integrase activity and

are responsible for the whole retrotransposition process. LINE1 is the youngest class and the only one that is still capable to retrotranspose.

Alu elements belong to the class of SINEs. *Alu*s are 300-bp repetitive sequences that are transcribed by Polymerase III (Pol III) from an internal A box and B box [9,10]. Even though *Alu*s are non-autonomous retrotransposons, these elements have been particularly efficient in exploiting LINE1 retrotransposition proteins, reaching a copy number of 1 million sequences (10.6% of the human genome), whereas LINE1 only account for 516˙000 copies [1].

This peculiarity, along with other features described in the next paragraphs, drew our attention on the study of *Alu* elements.

Structure of Alu elements

Alu repeats are thought to be originated from the retrotransposition of the 7SL RNA gene, giving rise to the dimeric structure of *Alu* elements. Indeed, the approximately 300-nucleotide sequence of *Alu* transposons consists of a left arm and a right arm. As already mentioned, the left arm harbours a type II internal Pol III promoter, composed of an A box and B box, with the right arm ending with a poly(A) tail. The two *Alu* monomers are separated by a characteristic A_5TACA_6 sequence (Figure 4). The region between the poly(A) tail and a canonical or non-canonical Pol III terminator (4 Ts or a T rich-region, respectively), is called 3' trailer and is unique to each individual *Alu* RNA. *Alu* elements are flanked by Target-site Duplications (TSD), which derive from the retrotransposition event.

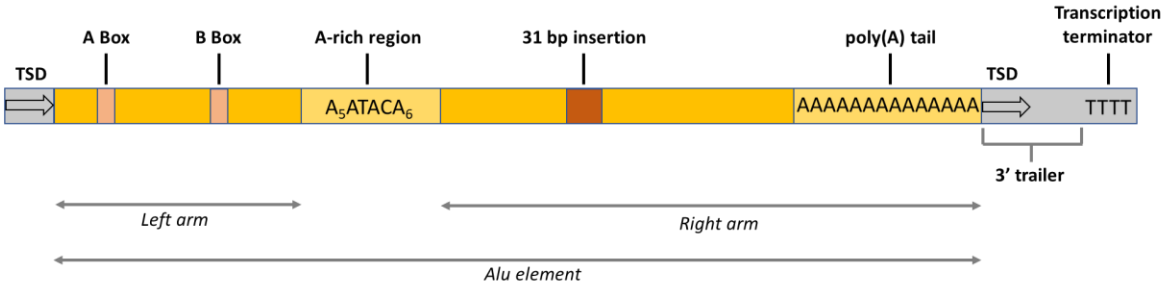


Figure 4. Structure of a canonical *Alu* element. The body of a typical *Alu* sequence is characterised by a left arm, which harbours the Pol III promoter, and a right arm, which ends with a long poly(A) stretch (25-100 nucleotides). The element is flanked by direct repeats (TSD) that originated during the retrotransposition process (TSD=Target Site Duplication). The 3' trailer region is specific of each *Alu* locus, extending from the end of the poly(A) tail to the first canonical or non-canonical Pol III terminator. Adapted from [11].

Retrotransposition Mechanism of LINE1

As already mentioned, *Alu*s are non-autonomous retrotransposons that completely depend on LINE1 (L1) elements for their re-insertion into the genome. Therefore, a brief discussion about L1 retrotransposition mechanism is necessary in order to clarify the extraordinary *Alu*s' efficiency to amplify in the human genome.

After transcription by Pol II, the L1 mRNA is exported into the cytoplasm, where it is translated in the ORF1p and ORF2p proteins. These two proteins bind the L1 mRNA with *cis* preference, according to a model where LINE1 RNA co-translationally binds nascent L1 proteins, forming an RNP complex that is the intermediate of retrotransposition [12] (Figure 5). ORF1p is an RNA binding protein with chaperone activity and was demonstrated to trimerize in order to bind the LINE1 mRNA [13], whereas ORF2p has endonuclease [14] and reverse transcription activity [15]. The efficient translation of ORF1p allows the ORF1p trimers to bind multiple times in order to coat the transcript [13,16,17], whereas ORF2p is present at only one or two copies at the L1 mRNA [18]. Once in the cytoplasm, the L1 RNP is re-imported into the nucleus,

where ORF2p mediates the LINE1 reinsertion into another genomic locus through a Target Primed Reverse Transcription (TPRT) mechanism [19] (Figure 5).

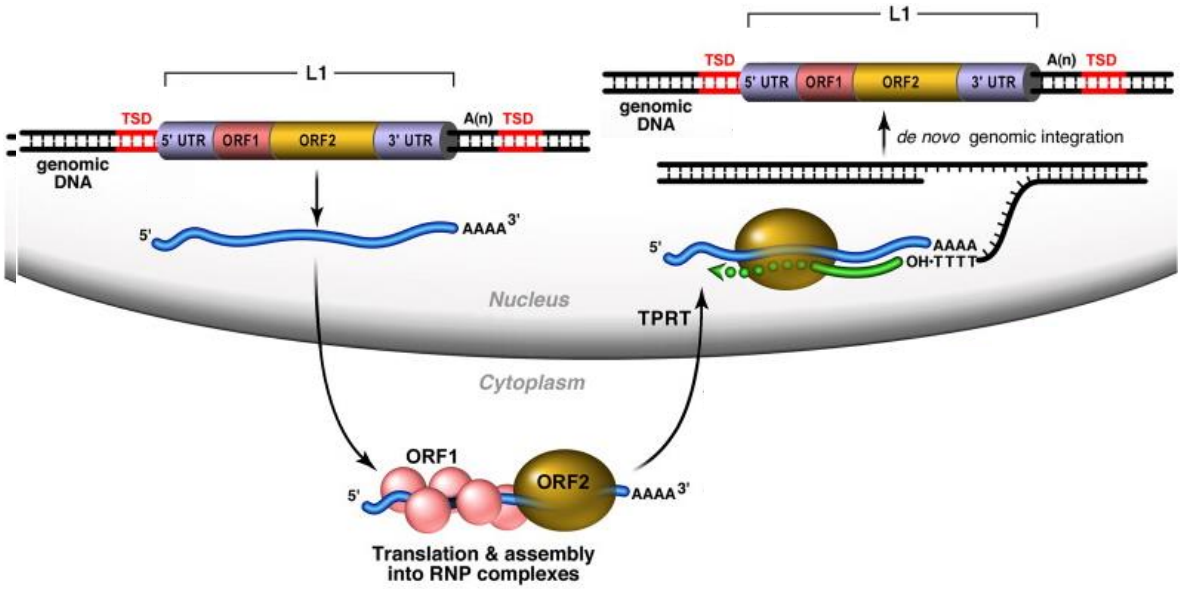


Figure 5. Retrotransposition mechanism of LINE1 elements. The LINE1 DNA sequence is transcribed and transported into the cytoplasm, where a ribonucleoprotein (RNP) complex is formed as a retrotransposition intermediate. The RNP complex is formed by the protein ORF1p and ORF2p that are translated from the same L1 RNA molecules. The association with the ORF proteins stabilizes L1 RNA and mediates the translocation into the nucleus. ORF2p creates a nick in the genomic target site, favouring the annealing between the LINE1 poly(A) and the genomic 3' OH end, from which the retrotranscription process can start. Adapted from [20].

A first nick is created at the 5'-AAAA/TT-3' consensus site generating a free 3'OH [21]. After the annealing of the poly(A) tail to the cleaved site, ORF2p can prime the synthesis of a cDNA using L1 mRNA as a template. A second nick is afterwards introduced in the genome, creating a second free 3'OH from which second strand synthesis occurs. At the end of the whole process, target site duplications (TSD) are created in the flanking regions of the L1 element (Figure 6).

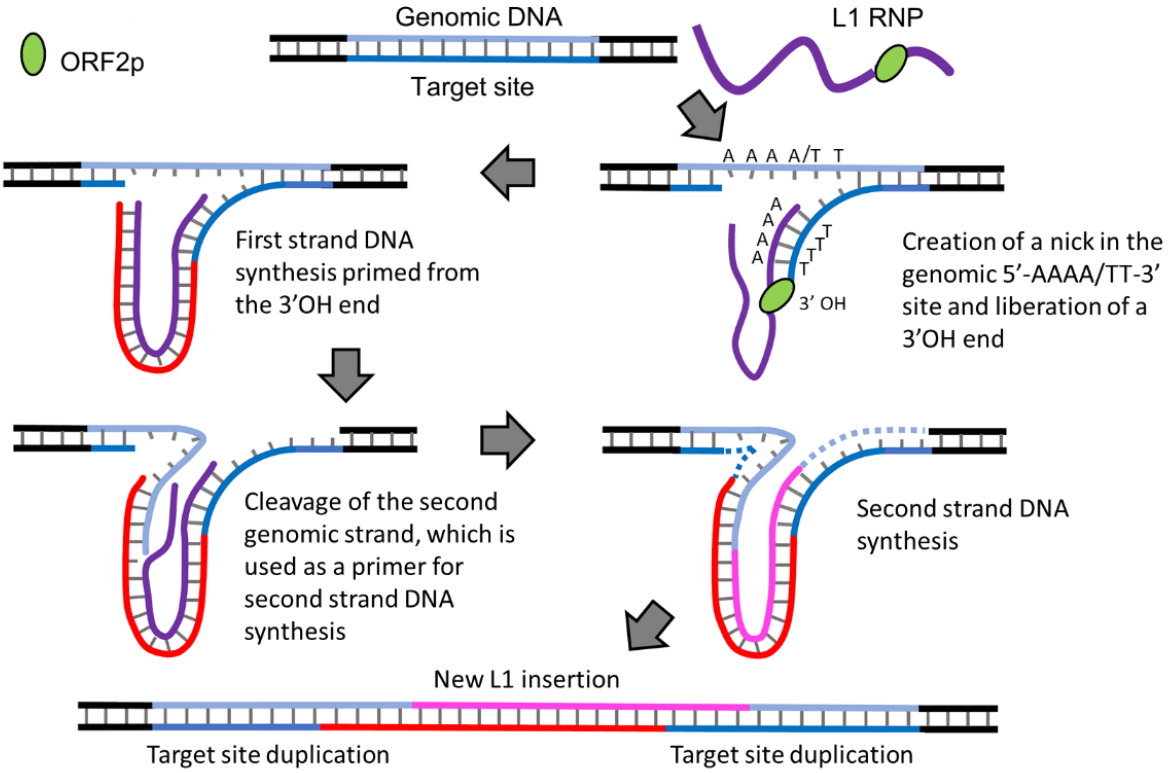


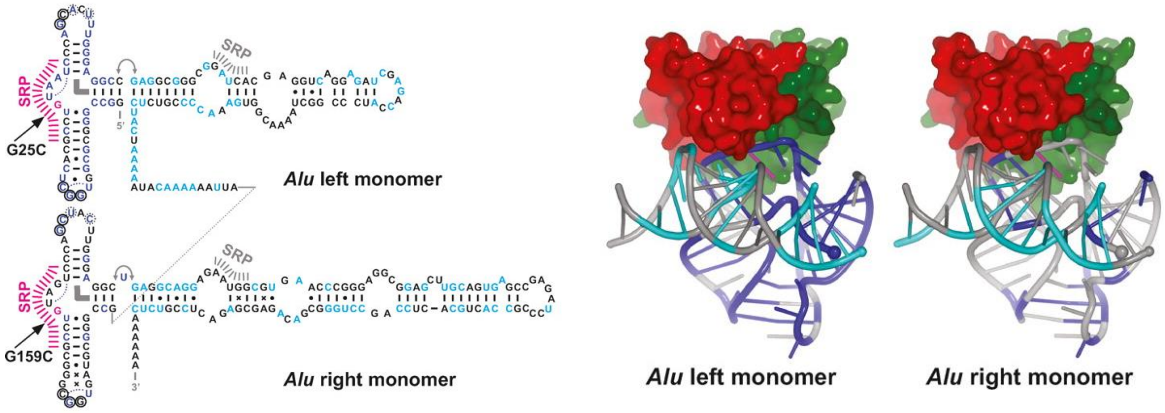
Figure 6. Mechanism of re-insertion of L1 elements into the genome. The first step of Target Primed Reverse Transcription (TPRT) consists in the creation of a nick in the genome at an AAAA/TT target site. The free genomic poly(T) stretch anneals to the L1 poly(A) tail, priming the reverse transcription reaction mediated by ORF2p. After the creation of a second nick at the genome target site, the second strand of DNA is synthesized by ORF2p. At the end of the process TSDs are created upstream and downstream the new L1 element. Adapted from [22]

Retrotransposition of Alu elements

Alu retrotransposition is a relatively efficient process, since it is estimated to happen every 20 human births [23]. It was estimated that 852 *Alu* sequences are still capable to retrotranspose [24], whereas only 80-100 copies of retrotransposition-competent L1s are thought to exist in the human genome [25].

It is outstanding how *Alu* elements have been able to hijack LINE1 retrotransposition for their own amplification, and various factors such as a characteristic nucleotide structure and RNA folding are thought to have a major influence in the whole process.

As already mentioned, the two *Alu* monomers are derived from the 7SL gene, which is part of the Signal Recognition Particle (SRP). Each *Alu* monomer, therefore, is able to bind the heterodimer SRP9/14, forming an *Alu* SRP complex (Figure 7) [14,24].



(A) (B)

Figure 7. Schematic representations of the interactions of *Alu* RNA with the SRP9/14 heterodimer. (A) Secondary structure of the *AluYa5* RNA. Hash marks indicate SRP contact sites. (B) Representation of *Alu* RNA binding to the SRP9/14 heterodimer. SRP9 is in red and SRP14 is in green. The representation is based on the crystal structure of 7SL bound to SRP9/14, as reported in Weichenrieder et al., 2000 [14]. Adapted from [24].

Since the SRP9/14 heterodimer interacts with the ribosome to halt protein synthesis and to direct the translation system to the endoplasmic reticulum, it is thought that *Alu* SRP is also able to associate with the ribosome that is translating LINE1 ORFs. This would bring to an efficient capture of ORF2p by *Alu* RNA [26].

Moreover, the association of *Alu* RNA to ORF2p is also promoted by the poly(A) binding protein (PABP). The PABP binds to approximately 25-adenosine stretch [27] and is thought to assemble on *Alu* poly(A) tail in a cooperative manner. A longer poly(A) tail would therefore allow the binding of a higher number of PABP molecules, which can also interact with the LINE1 CAP to stabilize the association of *Alu* SRP with LINE1 RNP [28,29] (Figure 8).

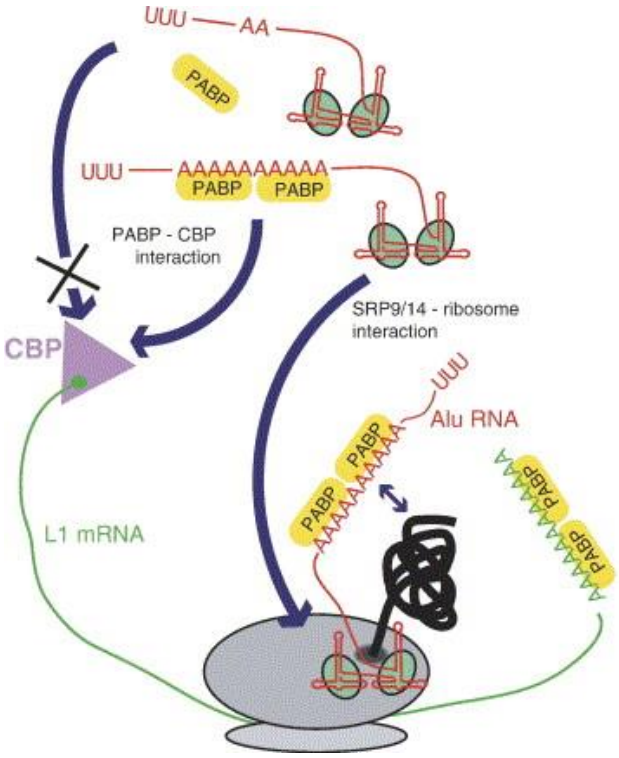


Figure 8. *Alu* RNA hijacks the LINE1 retrotransposition system. The *Alu* SRP complex is thought to interact with the ribosome engaged in the process of LINE1 translation, facilitating the association of *Alu* SRP to ORF2p. The poly(A) tail of *Alu* elements is thought to have an important role in the retrotransposition process. The poly(A) stretch can be recognized by the poly(A) binding protein, which in turns interacts with ORF2p and LINE1 mRNA CAP, stabilizing *Alu* RNA-LINE1 RNP interaction. Taken from [28].

Once in the nucleus, *Alu* re-insertion in the genome takes place through a TPRT mechanism as in the case of LINEs. Because *Alu* elements do not encode their own Pol III terminator, *Alu* RNAs contain a unique 3′ sequence that extends from the last nucleotide of the poly(A) tail to the first canonical or non-canonical Pol III terminator (a stretch of four Ts or a T-rich terminator), which is found downstream of the new insertion genomic locus. Since the reverse transcription is primed from the poly(A) tail, the 3′ trailer associated with the parent *Alu* is lost, and a new 3′ trailer is acquired during the new retrotransposition event. The 3′ trailer is therefore unique in length and sequence to each *Alu* element [30-32].

Alu retrotransposition could be influenced by a multitude of factors. First of all, the rate of transcription is certainly a crucial aspect. As extensively discussed in the next paragraphs, *Alu* RNA synthesis is a very limited process that can be influenced by the preservation of an intact Pol III promoter [33], the presence of flanking sequences that can stimulate Pol III transcription [34-36], and by the deposition of epigenetic marks.

Secondly, DNA mutations could take place in positions that are crucial for a proper RNA folding and interaction with the SRP9/14 heterodimer. Indeed, studies in the last decades demonstrate a relationship between the amount of *Alu* sequence variation and the efficiency of transposition, with some mutations being more efficient than others in affecting the level of activity [24]. Therefore, the presence of an intact body sequence might influence the efficiency of the retrotransposition event [37].

Third, the importance of the length of the poly(A) tail is suggested by the observation that old, retrotranspositionally not active *Alu* families have a mean length of 21-26 adenines, whereas the poly(A) length of more recently inserted *Alu*s ranges from 40 to 97 nucleotides [29]. This indicates that the poly(A) tail is prone to shrink after a new *Alu* insertion into the genome, possibly due to a strand slippage during DNA replication [38]. However, it was shown that *Alu* elements with a long tail are not advantaged in the retrotransposition process, since the transposition efficiency decreased for elements longer than 50 nucleotides [28]. Moreover, the preservation of a homogeneous tail seems to be another factor that influences the retrotransposition efficiency, since older elements tend to have a more heterogeneous tail than younger ones [32].

Even though *Alu* retrotransposition completely depends on LINE ORFs, there are numerous differences between the genomic distribution of *Alus* and LINEs. The presence of a higher *Alu* copy number compared to LINEs could be explained by the requirement of ORF2p but not ORF1p for *Alu* retrotransposition. Therefore, LINE1 copies that express only ORF2p (possibly deriving from splice variants, from mutated ORF1p or from a truncated LINE1) can still support *Alu* retrotransposition but would be defective in promoting their own retrotransposition [39,40]. Moreover, probably due to *Alu*s' shorter sequence compared to LINEs, it is thought that *Alu* retrontrasposition takes place faster than LINE1's [41].

Interestingly, even though *Alu* sequences depend on the TPRT mechanism, preferring therefore AT rich sequences for their reinsertion into the genome as in the case of

LINEs [1,42], *Alu* elements are most often found at GC-rich genomic sequences [1]. It is reasonable to think that *Alu* insertion initially takes place at AT-rich sequences, but the genomic distribution is subsequently reshaped by evolutionary forces. Indeed, evolutionarily old *Alu* sequences show a preference for GC-regions, whereas young *Alu* elements are instead more often found in AT-rich sequences [1]. This suggests a positive selection of *Alu*s in GC-rich regions, leading to hypothesize that the retention of *Alu* sequences would bring a benefit for the organism.

Evolution of Alu sequences and Alu subfamilies

It is well known that the 7SL RNA gene is particularly suitable for retrotransposition, as supported by the presence of several hundred retro-pseudogenes in mammalian genomes [43]. Therefore, it is not hard to believe that *Alu* sequences originated from ancestral monomers derived from 7SL [44]. Several studies suggest that this retrotransposition event gave rise to two lineages. One lineage, restricted to primates, generated the fossil *Alu* monomer (FAM), which subsequently evolved in a Free Right *Alu* monomer (FRAM). A second lineage, common to the whole supraprimates clade, is thought to be evolved in the FLAM-C subfamily of the free left *Alu* monomer [45] (Figure 9). The subsequent head-to-tail fusion of the FLAM-C and FRAM elements gave rise to *Alu* sequences about 65 Mya [46], starting an extensive process of *Alu* amplification in the human genome.

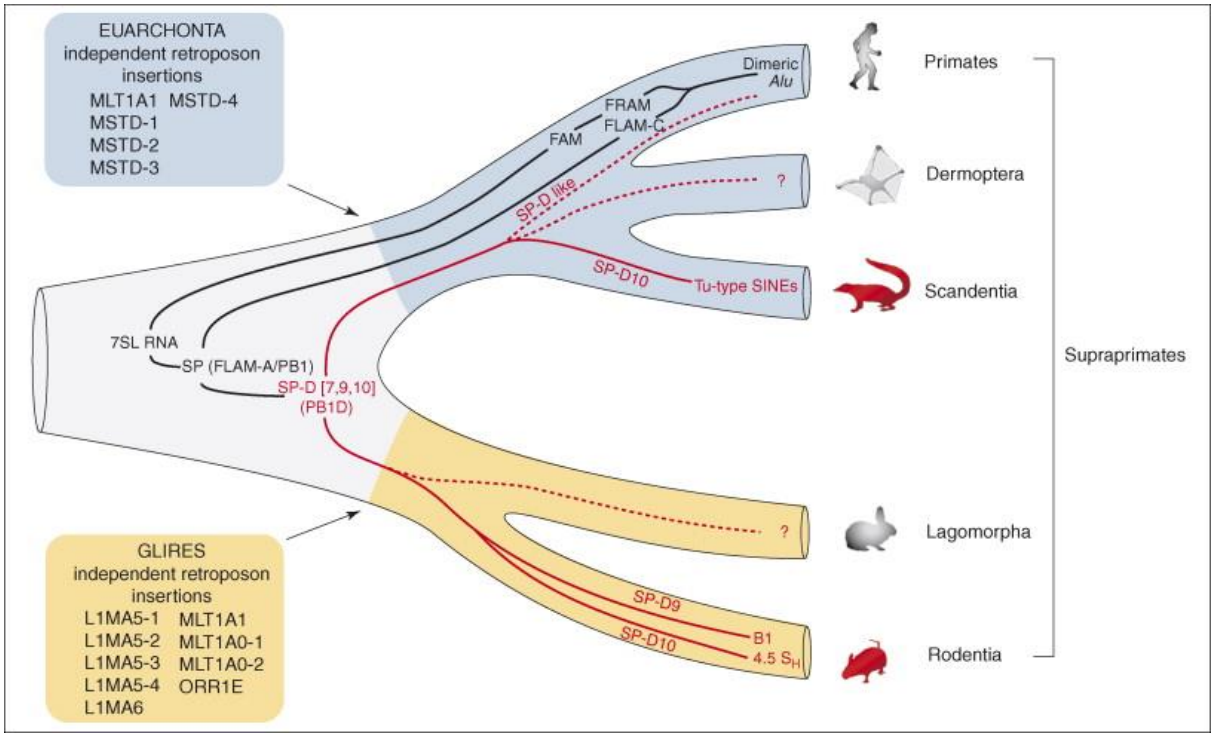


Figure 9. Evolution of *Alu* sequences. The retrotransposition of the 7SL RNA gene gave rise to the FAM element, which evolved in the FRAM sequence (Free Right *Alu* Monomer) and to the FLAM-C element (Free Left *Alu* Monomer). The subsequent head-to-tail fusion of the FLAM and FRAM elements gave rise to *Alu* retrotransposons 65 Million Years Ago. Taken from [45].

*Alu*s expanded from a limited number of master sequences with different rates of amplifications through the evolution, leading to the formation of three main subfamilies: *Alu J*, *Alu S* and *Alu Y*. The most ancient sequences belong to the *Alu J* subfamily, which is present in all primates and reached roughly 160˙000 copies in the human genome. *Alu*s had a burst of expansion around 55 Mya, leading to the evolution of the *Alu S* subfamily [47]. Nowadays, *Alu S* are the most abundant sequences in the human genome, being present at about 650˙000 copies. The dominant subclasses are *Alu Sx*, *Alu Sq*, *Alu Sc* and *Alu Sp*. The *Alu S* subfamily gave rise to the most recent class of *Alu Y*, with *AluYa5* and *AluYb8* being the most retrotranspositionally active subclasses (Figure 10).

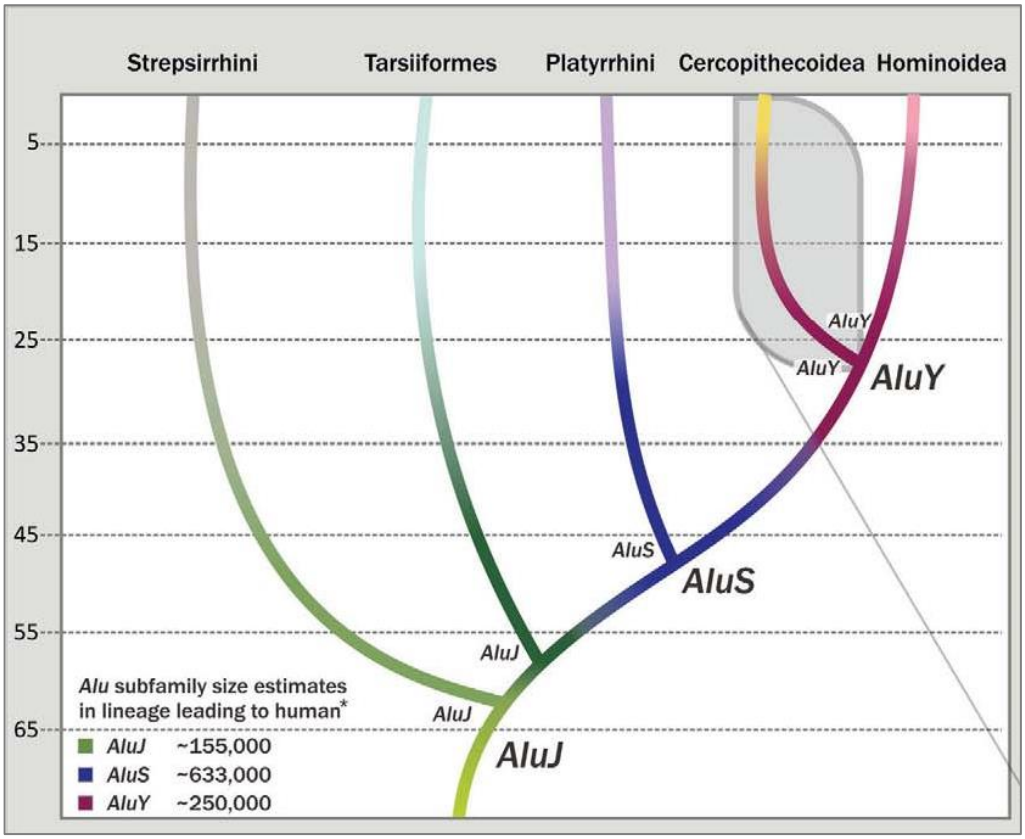


Figure 10. Evolution of *Alu* subfamilies in primates. *Alus* amplification started 65 Mya from *Alu J* subfamily. *Alu* amplification reached a peak with *Alu S* subfamily, which originated 55 Mya and led to the existence of roughly 650˙000 copies in the human genome. *Alu Y* originated from *Alu S* about 55 Mya and is the most retrotranspositionally active sequence in the human genome. Taken from [47]

During the evolution, *Alu* sequences have accumulated mutations that resulted in the inactivation of their retrotransposition. Indeed, it was calculated that the average *Alu J* copy accumulated 52 changes, making *Alu J* sequences not capable to mobilize to different loci in the genome [24]. By contrast, it is thought that only a few elements belonging to the *Alu S* family are still capable to retrotranspose, whereas the highest level of retrotransposition activity is reported for *Alu Y* sequences [24].

Distribution of Alu elements in the human genome

The reinsertion of transposon elements in the genome is probably favoured by relaxed chromatin conformations, which are more often found nearby actively transcribed genes. Indeed, *Alu* sequences are enriched in gene-rich regions [1], where they could be inserted both at intergenic or intragenic loci, and both in a sense or antisense orientation compared to the nearby coding gene (Figure 11). Moreover, intragenic *Alu*s could be present in any region of the gene: UTRs (Untranslated Regions), exons or introns [31]. Importantly, intergenic or antisense *Alu*s are expected to be exclusively transcribed by Pol III, whereas intragenic or sense *Alu*s could also be transcribed by Pol II as sequences embedded in protein coding transcripts [31]. It is important to be able to discriminate *Alu*s transcribed by Pol II than *Alu*s transcribed by Pol III, as it will be discussed in the next Chapter.

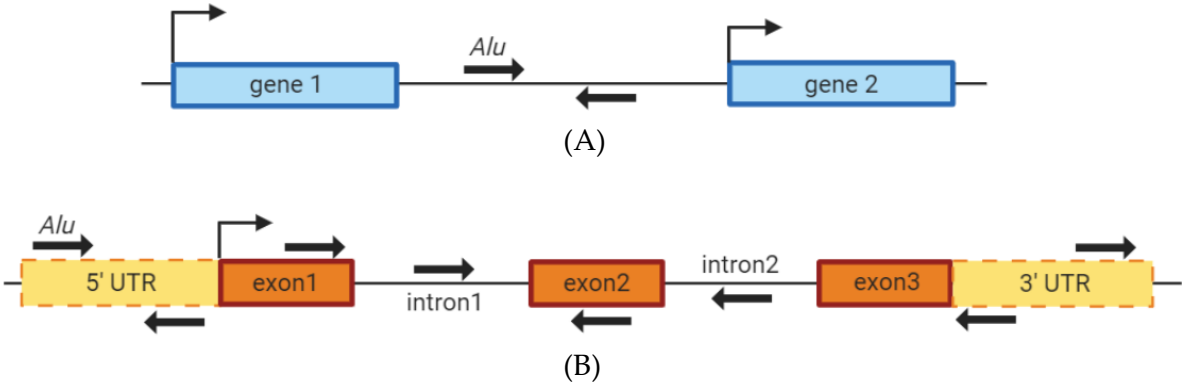


Figure 11. Genomic location of *Alu* elements. *Alu* elements can be inserted in intergenic regions (A) or intragenic loci (B). In the second case, *Alus* can be located in 5′ UTRs, 3′ UTRs, exons or introns. In all cases (intergenic or intragenic) *Alu* elements could be inserted both in a sense or antisense orientation compared to the protein-coding gene.

Transcription of Alu elements by Polymerase III

As already mentioned above, genuine *Alu* elements (i. e. not embedded in longer Pol II transcripts) are transcribed by Pol III from a type 2 promoter. The A and B boxes are first recognized by a multi-subunit complex known as TFIIIC, which then recruits the transcription factor TFIIIB. TFIIIB is composed of three subunits: the TATA box binding protein (TBP), BDP1 and the TFIIB related factor BRF1. TFIIIB is the main transcription factor that is involved in the recruitment of Pol III (Figure 12), even though protein-protein interactions between Pol III and TFIIIC may also contribute to stabilize the Polymerase III transcription machinery [48].

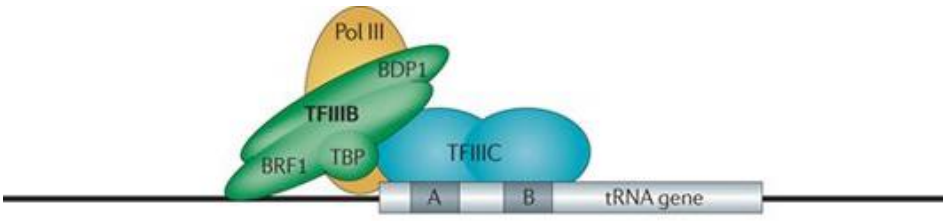


Figure 12. Assembly of the Polymerase III machinery to the tRNA-like *Alu* promoter. The first event in the assembly of the Pol III transcription machinery is the binding of TFIIIC to the type 2 promoter (as in the case of tRNA genes). TFIIIC is afterwards recognized by TFIIIB, which is responsible to recruit Polymerase III to the target gene. Taken from [49].

Pol III transcription is known to be cell line- and tissue-specific, as reported for tRNA genes [49]. Indeed, *Alu* transcription is a highly variable process that depends on the type of cell line or tissue [31,35,50], on growth conditions and on cell transformation state [51]. This transcription variability and the very low synthesis of *Alu* RNAs constitute a high challenge for *Alu* expression profiling. The low level of *Alu* transcription is due to a combination of low Pol III promoter strength in recruiting the

Pol III machinery [52] and a strong epigenetic silencing. An extensive discussion on *Alu*s' epigenetic repression will be included in the next paragraphs.

Recent works have shown that *Alu*s are inefficient in driving autonomous transcription [52], therefore the presence of flanking sequences is often required to enhance their expression [34,36,53].

Alu transcription was estimated to take place from only 100 loci in HeLa cells, giving rise to only very few copies of *Alu* RNA [35,54]. The low expression of *Alu*s in HeLa cells is further supported by the findings that only 13 loci were associated with the Pol III transcription machinery [55]. A low level of Pol III *Alu* transcripts is also observed in other types of tumor cells, embryonic stem cells and primary cells [31,54], and this is further confirmed by the finding that only 162 *Alu* loci were bound by Pol III factors in a total of six different cell lines [51].

Pol III transcription is regulated by many transcription factors that are also involved in Pol II regulation. For example, the oncosuppressor Rb is known to inhibit Pol III transcription through binding with Brf1-TFIIIB, hindering the formation of the preinitiation complex and therefore Pol III recruitment [56]. On the counter side, TFIIIB association with c-Myc is known to mediate the activation of tRNA genes, along with interactions with the cofactor TRRAP and the histone acetyltransferase GCN5 [57].

Pol III transcription of *Alu* sequences can be enhanced by the AP-1 transcription factor [34] and its subunits JUND [31]. Moreover, the CCCTC-binding factor (CTCF) and the CCAAT/Enhancer Binding Protein β (CEBP) were found enriched at *Alu* loci in different cell lines [31], in support of the knowledge that Pol III transcription is a highly tissue- and cell line-specific process [49]. On the other hand, *Alu* transcription is repressed by the CGG triplet repeat-binding protein 1 (CGGBP1) in response to growth factors [58].

Alu epigenetic silencing: DNA methylation

Alu elements are highly enriched of the CpG dinucleotide, representing 25% of the CpG sites in the whole genome [59]. CpGs are highly vulnerable of methylation at the C residue, which is known to mediate transcriptional repression through two principle mechanisms. First, DNA methylation might prevent the binding of transcription factors to the gene promoter [60]. Second, methylated CpGs can be recognized by methyl-CpG-binding proteins (MBPs), which in turn interact with transcriptional repressor complexes.

A first indication of the involvement of DNA methylation in *Alu* transcriptional repression was supported from *in vitro* transcription assays: plasmid treatment with DNA methylases resulted in the repression of Pol III *Alu* transcription, possibly through the MeCP1 histone deacetylase complex [33]. *In vivo* studies further confirmed a transcriptional repressive role of DNA methylation, since Pol III *Alu* transcription was enhanced in HeLa cells treated with the DNA methylation inhibitor 5-azacytidine [61].

Moreover, a link between DNA methylation and *Alu* transcription was also verified in colon and lung cancer cells, where highly expressed *Alu*s were found hypomethylated [62,63]. However, the technique used for the detection of *Alu* RNAs did not allow to discriminate between a genuine Pol III *Alu* or an mRNA embedded *Alu*. Therefore, the increase in *Alu* followed by hypomethylation could also be caused by a de-repression in Pol II transcription.

Alu histone modifications

Other than DNA methylation, histone modification was found associated with *Alu* silencing, supporting an alternative mechanism for *Alu* transcriptional repression. In contrast with Liu et al. [61], the treatment of HeLa cells with the DNA methylation inhibitor 5-azacytdine was demonstrated to have no impact on *Alu* transcription,

whereas the trimethylation of H3K9 would play a major role in *Alu* repression [64]. Indeed, Pol III occupancy at *Alu* loci was not altered by 5-azacytdine, whereas it was enhanced after the inhibition of the histone methyltransferase SUV39H1 leading to a consequent increase of *Alu* RNA. The epigenetic silencing of *Alu* transcription would therefore be caused by H3K9me3 instead of DNA methylation.

H3K9me3 is only one of the various epigenetic marks involved in the regulation of *Alu* transcription. Indeed, several works show that *Alu* sequences are enriched of histone modifications that are characteristic of enhancer elements, which were recently found to be actively transcribed [65]. Therefore, it is tempting to hypothesize that the enrichment of enhancer marks at *Alu* elements could also influence *Alu* transcription. Enhancers are identified by (i) the presence of DNA motifs responsible to recruit Transcription Factors (TFs), (ii) the hypersensitivity to DNAse treatment, which is characteristic of a relaxed chromatin state, (iii) histone marks that are indicative of transcriptional activation (high H3K4me1/H3K4me3 ratio and/or H3K27ac) or transcriptional repression (H3K27me3 and H3K9me3), (iv) the presence of histone variants such as H2AZ, (v) the binding of transcriptional co-factors, such the histone acetyltransferase p300, (vi) the ability to influence genes that are distant 10^3-10^6 bases by chromatin looping [66]. *Alu* elements are characterized by many of the enhancer features that are listed above. Indeed, in a very recent study it was demonstrated that Pol III *Alu* transcripts had signatures of enhancer marks in a cell line-dependent pattern, bearing sensibility to DNAse digestion, enrichment in the histone variant H2AZ1 and in the histone activating marks H3K4me1/2, H3K27ac and H3K9ac [50]. H3K4me1 enrichment at *Alu* elements is also confirmed in CD4+ T cells and EBV-transformed lymphocytes at intergenic *Alu*s, where Pol III transcription is more likely to occur than Pol II transcription [67]. Moreover, the coexistence of Pol III and the activating epigenetic mark H3K4me3 was found at *Alu Y* elements in human testes, indicating a positive role of H3K4me3 in Pol III *Alu* transcription [68].

Alu activation is mediated by the AP-1 protein complex, which was found associated to Pol III *Alu* loci together with p300 and H3K27ac [50]. Moreover, it has to be noted that H3K9me3 was not found enriched in Pol III expressed *Alu*s, confirming the repressive role of H3K9me3 in *Alu* transcription [64,69].

In conclusion, DNA methylation and H3K9me3 seem to act as repressive marks for *Alu* transcription, whereas H3K4me1/2/3, H3K27ac and p300 are likely to be involved in facilitating *Alu* expression. Lastly, in addition to the hypothesized contribution of transcription factors (as mentioned in the previous paragraphs) and histone modifications, *Alu*s' transcriptional increase is also observed after cell perturbations that induce a global chromatin remodeling. Indeed, *Alu* overexpression was detected in senescent cells [70] and in response to cell stresses such as heat shock [71-73] and virus infection.

Activation of Alu and other SINEs by viral proteins

Pol III transcription of *Alu* sequences, and more generally of host and virus Pol III genes, can be enhanced by virus infection.

The Simian Virus 40 (SV-40) increases the transcription of B2 SINE loci in mouse cell lines [74]. The large T antigen of SV-40 is thought to play a crucial role in the activation of Pol III transcription increasing TFIIIB and TFIIIC expression and releasing the repressive interaction between Rb and TFIIIB, as shown in rat cell lines [75].

B2 SINE transcription can also be enhanced during murine gammaherpesvirus 68 (MHV68) infection [76]. In this case, the observed transcriptional increase is probably modulated by a more complicated molecular mechanism. Indeed, SINE profiling through RNA-seq revealed that only a subset of sequences is activated in response to MHV68 infection, and this activation seems to be independent from the integrity of the SINE promoter [77].

Importantly, virus-induced SINE transcription has also been detected in different human cell lines in response to Herpes Simplex Virus (HSV) and Adenovirus infections. HSV increases the transcription of *Alu* repeats through stimulation of the TFIIIC activity operated by the HSV immediate-early protein ICP27 [78].

Moreover, *Alu* expression is enhanced in HeLa and embryonic kidney cells by Adenovirus 2 and Adenovirus 5 infections [71,79-82], whereas *Alu* activation in human fibroblasts was reported to be increased only by Adenovirus 5 infections [82]. The molecular mechanism responsible of Adenovirus-induced *Alu* activation is mostly unknown. However, several studies reported that the early E1A protein is able to induce the transcription of the Pol III viral genes VAI and VAII influencing TFIIIC DNA-binding properties [83], and therefore promoting TFIIIC transcriptional activity [84,85]. In addition to the knowledge that E1A can induce host global epigenetic changes [86-89], it is tempting to hypothesize its involvement in *Alu* activation, even though a direct role has never been experimentally demonstrated. Further discussion about E1A-induced epigenetic reprogramming will be included in the next paragraphs.

Life cycle of Adenovirus 5

Alu transcription is enhanced by Human Adenovirus 5 (Ad5), even though it is not clear which virus factors are involved in this process. However, it is reasonable to think that viral proteins that interact with chromatin remodelers could have a major impact on *Alu* transcription. A brief discussion about Ad5 life cycle would help to contextualize the roles of viral proteins in the activation of *Alu* sequences.

Adenovirus 5 is a DNA virus that infects terminally differentiated cells in the respiratory tract. Quiescent cells are not a proliferative environment for virus replication, therefore Ad5 has to force cell cycling in order to drive the synthesis of its own DNA. Ad5 genome is a linear double-stranded DNA of 36-kb and contains five

early transcription units (E1A, E1B, E2, E3 and E4), four intermediate transcription units (IX, IVa2, L4 intermediate and E2 late) and one late transcription unit that gives rise to 5 classes of late mRNAs (L1, L2, L3, L4, L5) (Figure 13).

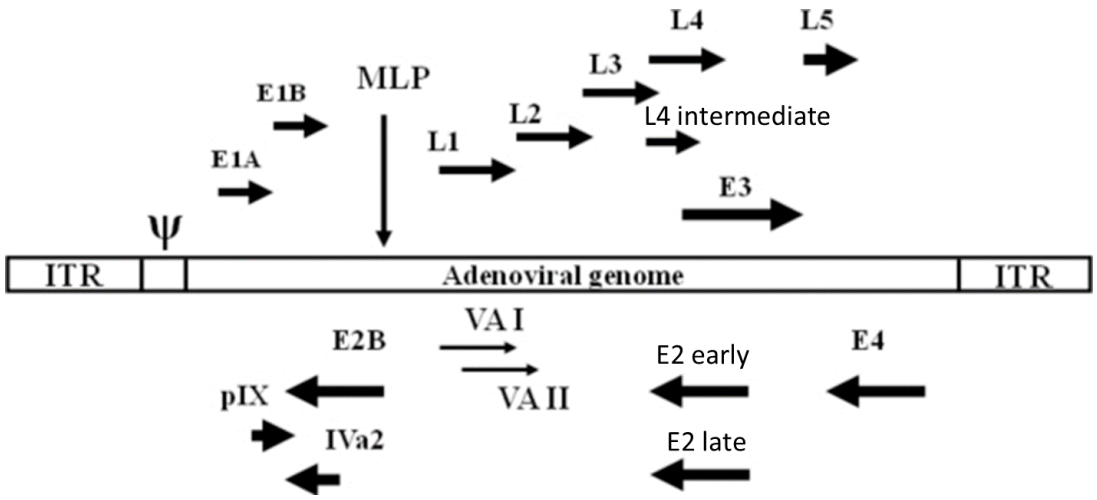


Figure 13. Adenovirus type 5 genome. Early genes correspond to E1, E1B, E2, E3, E4; intermediate genes are indicated as IX, IVa2, L4 intermediate, E2 late. Late genes are marked as L1-L5. These genes are transcribed by Pol II, whereas the viral genes VAI and VAII are transcribed by Pol III. MLP represents the major late promoter, which drives transcription of the late genes. The 36-kb Ad5 genome is flanked by inverted terminal repeats (ITR). Ψ indicates the packaging site. Adapted from [90].

All viral genes are transcribed by Pol II except for VAI and VAII, which are transcribed by Polymerase III.

Immediately after infection, the Ad5 genome is translocated into the nucleus, where E1A is the first gene to be transcribed due to strong enhancer activity approximately 500-bp upstream its promoter.

E1A is an essential protein in the viral life cycle, since it activates its own transcription and the transcription of the other early and late viral genes. Therefore, Ad5 completely relies on E1A to conclude the viral life cycle. The proteins synthesized in the early phase of infection set up a prolific environment for viral replication: E3 inhibits cell

immune response to viral infection; E1B inhibits the p53-dependent induction of apoptosis through binding to the proapoptotic BCL-2 family members BAK and BAX and through binding the tumor suppressor p53; E4 is involved in the activation of mTOR signalling without the requirement of external nutrients, leading to an enhancement of protein synthesis. Moreover, E4 binds to the cellular transcription factor E2F, increasing its affinity for the promoter of the early E2 gene, which is therefore the last gene to be transcribed in the early phase of infection. E2 encodes three viral proteins required for DNA viral replication. After the beginning of viral DNA synthesis, the major late promoter (MLP) reaches a peak of activity, driving the synthesis of one single late transcription unit. This transcript is spliced at different poly(A) sites producing the 5 late mRNAs (L1-L5), which are involved in the regulation of translation during the late phase and in virion assembly. It seems that MLP activity is enhanced by changes in the viral chromosome that depend on viral DNA replication, and the requirement of the IVa2 and L4 proteins is also hypothesized. During the late phase of infection, large amounts of virion structural proteins are synthesized to generate ≈ 100˙000 virions per HeLa cells [91].

E1A interaction with host proteins

As mentioned in the previous paragraph, E1A is the first protein to be transcribed immediately after infection, and its essential role in viral life cycle consists in the activation of viral proteins of the early, intermediate and late phase.

Two E1A mRNAs are encoded from alternative splicing that uses one of two 5' splice sites, separated by 138-bp, and the same 3' splice site. The two alternative transcripts code for a 289 residues protein, called large E1A, and a 243 amino acid protein, called small e1a (Figure 14).

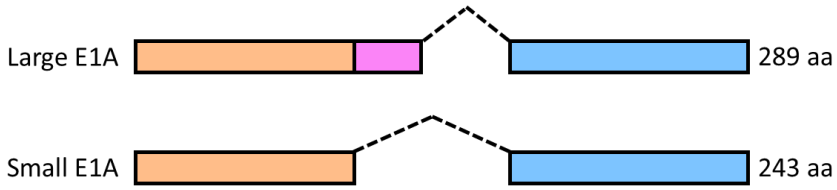


Figure 14. Schematic structure of E1A mRNAs. The alternative splicing of the same E1A transcript produces two E1A variants: large E1A codes for a 289 aminoacids (aa) protein, whereas small E1A is a 243 aa protein. The central domain of large E1A, highlighted in purple, is responsible for the interaction with the MED23 mediator subunit.

The ability of large E1A to induce transcription of the early viral genes relies on the interaction of its unique central domain with the Mediator of Pol II transcription subunit 23 (MED23). Small E1A (hereafter simply called e1a) lacks the MED23-interaction domain and therefore is not able to induce the transcription of the other early viral genes and to induce viral replication.

Other than activating viral genes, E1A proteins have a central role in inducing the host cell to enter the S-phase of the cell cycle. Both large and small E1A play roles in this process, being able to interact with host chromatin remodelling factors and gene regulators. Importantly, e1a is sufficient and necessary to induce a transformed phenotype in growth arrested cells [86,92].

From amino acid sequence alignments of E1A proteins deriving from different human Adenovirus serotypes, it is possible to identify regions of conservations (CR1-CR4) that are important to establish E1A-host proteins binding. The oncosuppressor Rb, the host lysine acetylases p300/CBP and the SWI2/SNF2-related p400 proteins are thought to be key interactors to induce cell transformation, as will be discussed in the next paragraphs.

The conserved amino acids in CR1 and CR2 regions span positions 41-80 and 115-140, respectively (Figure 15) [93,94]. The CR3 region is only present in large E1A and interacts with the host mediator to activate the expression of viral genes (Figure 15). Lastly, even though not as extensively conserved as the CR regions, the N-terminal portion (aa 13-38) is also conserved among primate adenoviruses [93] and is thought to be required to drive cells into the cell cycle [95] (Figure 15).

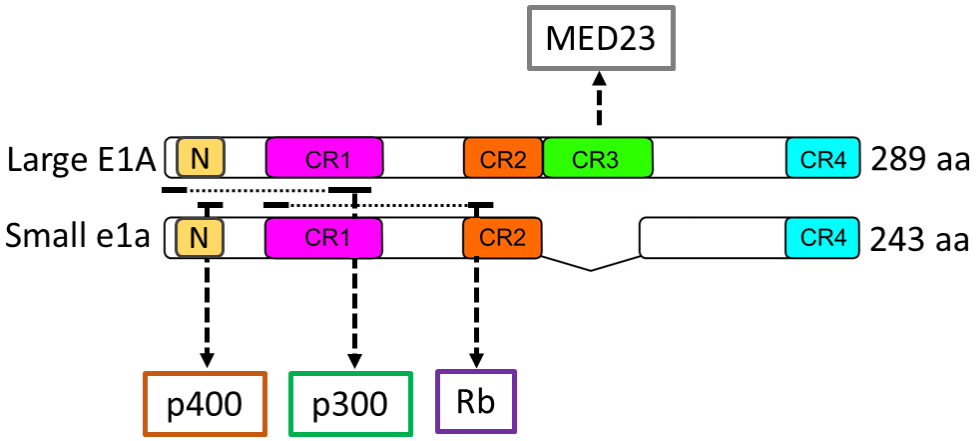


Figure 15. Schematic representation of large and small E1A structure and their interactions with the host proteins. The conserved regions CR1, CR2, and CR4 are common to both E1A variants, whereas CR3 is only present in large E1A. The N-terminal region is also characterized by a certain degree of conservation and binds the p300 and p400 proteins. The CR1 region mediates the interaction with p300 and Rb, which is also bound by five aminoacids in the CR2 region. Adapted from [96].

Oncogenic transformation induced by E1A: Rb and p300

As stated above, e1a induces extensive epigenomic and transcriptomic changes in the host cell and is sufficient to drive quiescent fibroblasts into the S-phase [86,87]. Interestingly, e1a has opposite effects on host gene expression, leading to the activation of genes involved in cell cycling and DNA replication, and to the repression of genes involved in the immune response. Both CR1 and CR2 are involved in this

process, and molecular interactions with the Rb protein and with p300/CBP were identified as key factors to drive cell reprograming.

Both small and large E1A bind Rb through N-terminal residues of the conserved region 1 (CR1) (aa 43, 44, 46, 47) and through the LXCXE motif in CR2 (aa 122-126) [93] (Figure 15). The molecular mechanism underlying E1A-Rb interaction to drive oncogenesis is very well studied and the crystal structure of Rb bound to E1A has also been characterized [97]. E1A binds Rb in the same pocket where Rb binds the transcription factor E2F, consequently liberating E2F from RB-inhibition and inducing the synthesis of cell cycle and viral genes [86,87,97]. In normal cell cycle, Rb is phosphorylated by cyclin-CDK complexes that are activated in response to extracellular mitogens, removing Rb repression from E2F to allow entry into the S-phase. E1A is therefore able to bypass cyclin regulation dependent from external stimuli, driving arrested cells into the cell cycle (Figure 16).

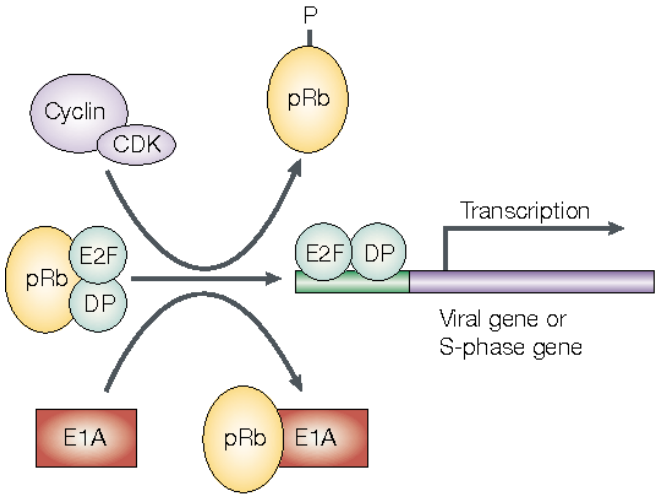


Figure 16. E1A interaction with Rb drives the host cell into the S-phase. In normal cycling cells, Rb inhibition of E2F/DP is released after Rb phosphorylation by cyclin-CDK complexes. E1A bypassed this regulation through binding to Rb, consequently liberating E2F and inducing the synthesis of host genes to enter into the S-phase. Adapted from [98].

Moreover, the increased availability of Rb proteins that are released from E2F transcription factors brings their accumulation at e1a-repressed genes involved in the cell immune response, development, differentiation and synthesis of the extracellular matrix, in a process that is independent from p300 binding [86,87].

E1A-p300 binding is essential to drive the reorganization of the host epigenome through a general redistribution of the histone acetylation patterns of the host cell.
E1A binds p300/CBP at amino acids 1-25 [99,100] belonging to the N-terminal region and amino acids 54-82 that belong to the CR1 region [100] (Figure 15). p300 is a transcriptional co-activator that is recruited to gene promoters through interactions with transcription factors [101]. Once recruited, p300 acetylates lysines located on histones and transcription factors through its intrinsic histone acetyl transferase activity (HAT), resulting in the enhancement of transcription. E1A binds p300 at its transcriptional adaptor motif domain [102], and it is thought that this interaction can inhibit p300 (HAT) activity [103], resulting in an inhibition and redistribution of p300 acetylation on different genomic loci and proteins.

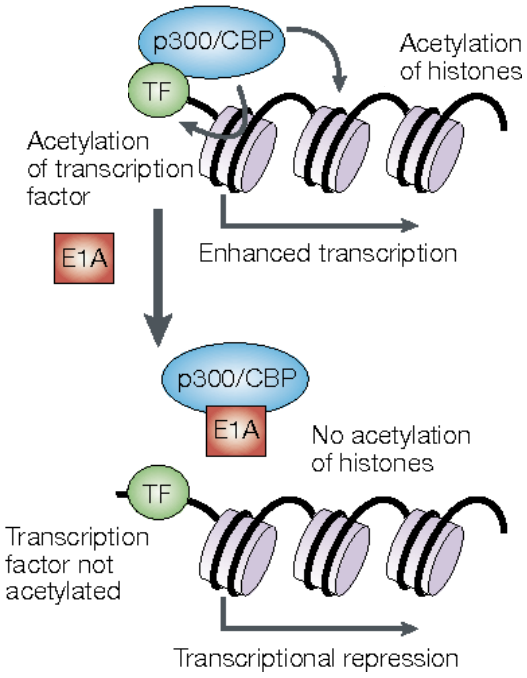
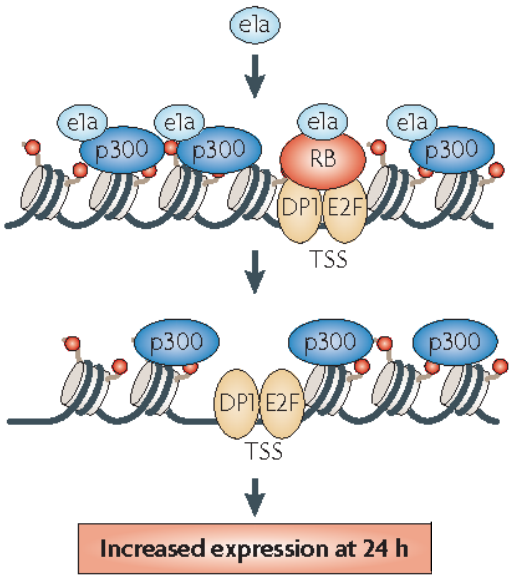


Figure 17. E1A inhibits p300 activity. p300 is a histone lysine acetylase that is recruited to gene promoters through interactions with transcription factors (TF). This results in the acetylation of histone tails and lysines of the TFs, consequently enhancing the transcription process. E1A prevents p300 activity by disrupting p300-TF interaction and by inhibiting p300 HAT activity, resulting in transcriptional repression. Adapted from [98].

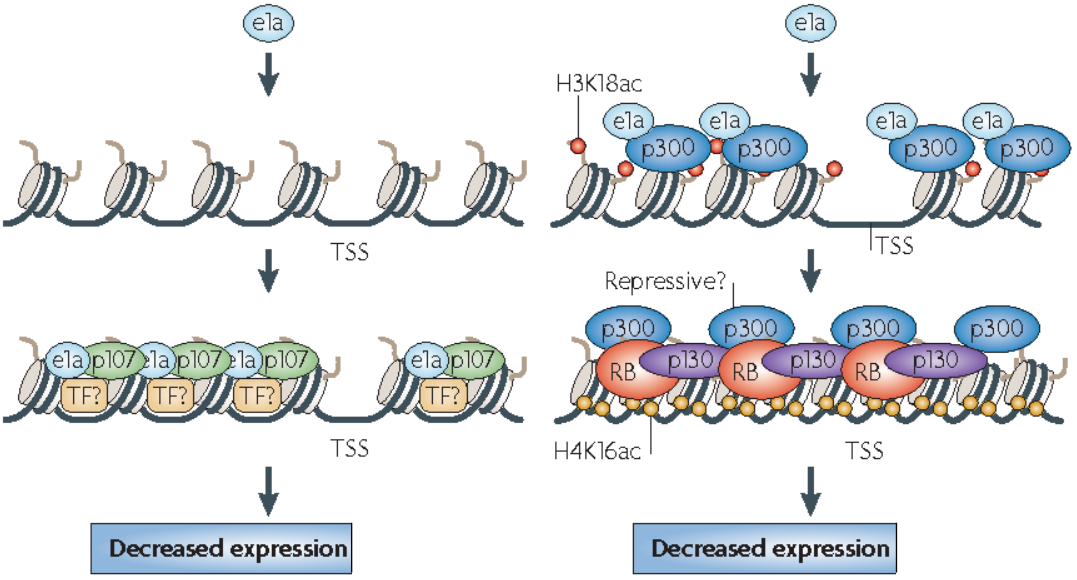
In accordance with the model described above, e1a interaction with p300 was demonstrated to induce a general decrease in H3K27ac [86], whereas H3K18ac was redistributed at activated genes and depleted at repressed genes. As predictable,

activated genes belong to cell cycle and DNA synthesis pathways [86-89]. Interestingly, the increase of H3K18ac at cell cycle gene promoters was also accompanied by a removal of Rb proteins, suggesting that the activation of cell proliferation genes driven by e1a involves the removal of an endogenous inhibition and the simultaneous deposition of an activating epigenetic mark by p300/CBP [86,88]. As seen above, a subset of genes is repressed by Rb independently than p300 binding, presumably exploiting the pool of Rb proteins that is dislocated from E2F. However, there is a subset of genes that seems to require the simultaneous binding of p300 and Rb to the same e1a molecule in order to be repressed. It is thought that in the p300-e1a-Rb trimolecular complex, e1a inhibits p300 HAT activity on H3K18 but enhances acetylation on the same e1a molecule and on Rb. H3K18 hypoacetylation, in conjunction with e1a and Rb acetylation, would induce chromatin condensation and gene repression [86]. In summary, it seems that e1a-induced gene activation would rely on Rb dislodging from E2F and the simultaneous histone acetylation. Gene repression results from Rb redistribution or chromatin condensation induced by the p300-e1a-Rb trimolecular complex, with the resulting decrease in hypoacetylation (Figure 18).



(A)

Figure 18. *Figure continued on next page*



(B)

Figure 18. Model explaining the reorganization of the host epigenome by small E1A (*continues from previous page*). e1a interacts with Rb oncosuppressor and the lysine acetylase p300, resulting in the activation of cell cycle genes and in the repression of a set of genes that hamper viral replication (A, from previous page). Gene activation is thought to be mediated by relocalization of H3K18ac on cell cycle genes and Rb dislodging from E2F. (B) Gene repression is presumably mediated by two alternative mechanisms. In a p300-independent mechanism, Rb (p107) is redistributed at repressed genes after dislodging from E2F. Other subset of genes could be repressed by the formation of a p300-e1a-Rb trimolecular complex, resulting in H3K18 hypoacetylation, H3K16 hyperacetylation and chromatin condensation.

Oncogenic transformation induced by E1A: p400

As already mentioned in the previous paragraphs, E1A binds several host proteins, among which the p400 protein was identified as essential target for E1A-induced cell transformation [105].

E1A mutants that lack a region spanning amino acids 26-35 are defective in binding a protein doublet of 400kDa (Figure 19), which was demonstrated to be constituted by the p400 protein and TRRAP (transactivation/transformation-domain-associated protein). p400 is homologous to the yeast SWI2/SNF2 nucleosome remodeling complex, as indicated by its ATP-dependent DNA helicase domain (Figure 19). E1A

interacts with p400 through a direct competitive binding with the TAP54 DNA helicase, whereas an indirect interaction occurs through the TRRAP protein (Figure 19).

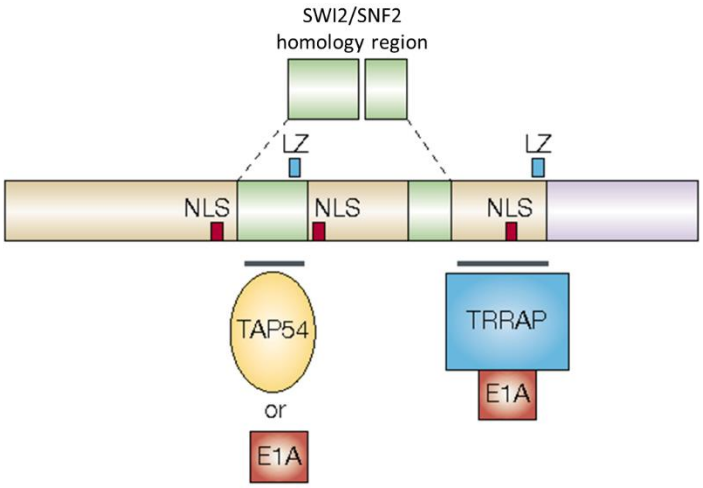
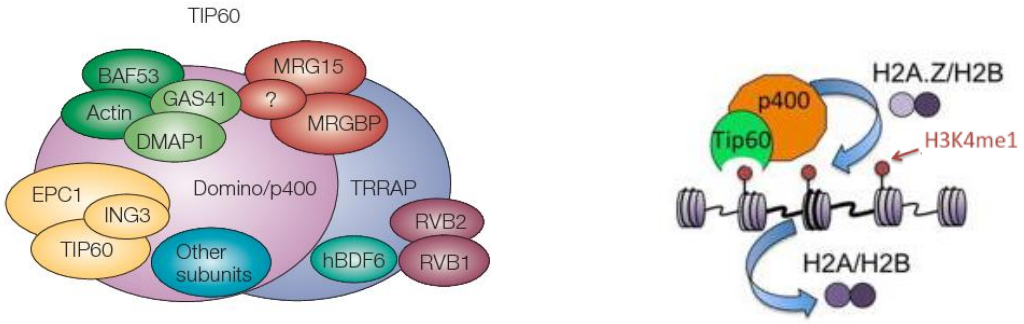


Figure 19. Schematic structure of the p400 protein. p400 interacts directly with E1A through a region that is also responsible to bind the DNA helicase TAP54, whereas an indirect binding is mediated by the transactivation/transformation-domain-associated protein (TRRAP). Adapted from [98].

Other protein subunits were also found associated in the same p400 complex copurified with E1A, such as the TAP54 DNA helicase (RVB2), the oncogenic transcription factor Myc, actin-like proteins and the human homolog of the Polycomb protein (EPC1), indicating a function of the p400 complex as chromatin remodeler [105].

Importantly, p400 and TRRAP are also found associated in a variety of cellular multiprotein complexes with HAT activity, among which the TIP60 complex is one of the most studied. TIP60 HAT activity is provided by the homonymous protein, and some common subunits of the p400 complex such as TAP54 helicases (RVB2) and the

homolog of the Polycomb protein (EPC1) are be part of the TIP60 complex (Figure 20A). The TIP60 complex is involved in a variety of cell functions, among which the enhancement of chromatin accessibility at enhancers has been reported. H3K4me1 recruits the TIP60 subunit through its chromodomain [106], favouring histone acetylation. Simultaneously, the TIP60 complex would promote nucleosome exchange through deposition of H2A.Z, which is also acetylated by TIP60 [107]. The whole process increases chromatin accessibility for transcription factors [108].



(A) (B)

Figure 20. Composition and cellular function of the TIP60 complex. (A) The TIP60 complex is composed of proteins that have ATPase activity (p400), HAT activity (TIP60), helicases (RVB1/2), DNA methyltransferase activity (DMAP1), as well as scaffold properties (TRRAP) and homology to the enhancer Polycomb1 (EPC1). Taken from [109]. (B) The TIP60 complex increases chromatin accessibility at enhancers. Tip60 is able to recognize H3K4me1 deposited at enhancer regions, leading to histone acetylation and nucleosome exchange mediated by p400. Adapted from [110].

The p400 complex and TIP60 complex are also shown to be involved in gene repression [111,112]. Interestingly, a recent work demonstrated that the TIP60 complex inhibits endogenous retroviral elements (ERVs) through stimulating the expression of methyltransferases SUV39H1 and SETDB and thereby H3K9me3 deposition [112].

The mechanism underlying cell transformation mediated by E1A-p400 interaction is not clear, although in the last years several hypotheses have been advanced.

e1a-p400 interaction could mediate gene repression in the context of e1a-Rb/p300 binding, forming a multiprotein complex that induces chromatin compaction [86].

On the other side, an involvement of E1A-p400 interaction in gene activation has been more extensively supported. For example, it seems that p400 acts in a similar manner as Rb to repress the E2F transcription factor. E1A-p400 interaction would displace p400 from E2F, releasing its transcriptional activity in a similar but independent way than E1A-Rb interaction [113].

Recent works suggest that the oncogenic transcription factor Myc is involved in cell transformation induced by E1A-p400/TRRAP interaction [114-118]. Indeed, some studies suggest that a direct binding between E1A and p400 induces Myc stabilization and its subsequent transcriptional activation. Indeed, E1A might promote the association of p400 with Myc, resulting in the inhibition of its ubiquitination. Importantly, the stabilized Myc protein could drive the p400 complex to Myc target sites on the DNA and activate Myc target genes [117,118] (Figure 21).

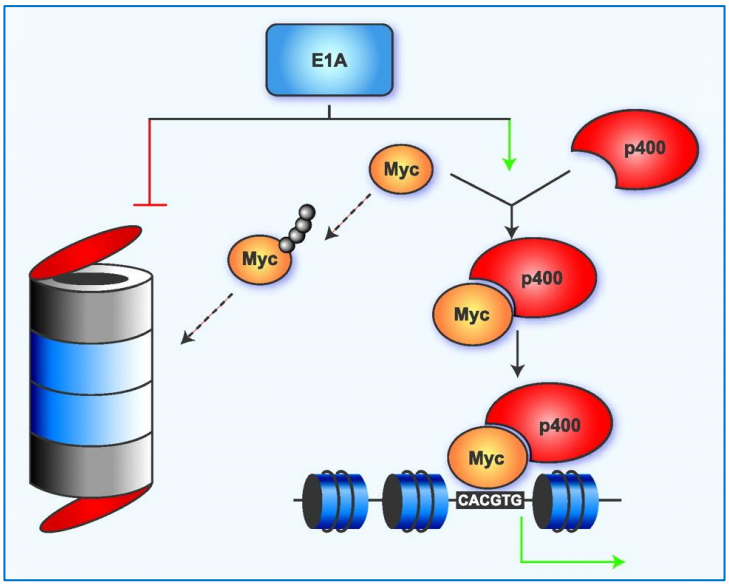


Figure 21. Model of Myc regulation by E1A. According to Chakraborty et al. [117], E1A recruits Myc at the p400 complex, where Myc ubiquitination is inhibited. Moreover, Myc may direct the p400 complex to its target genes, mediating transcriptional activation. Taken from [117].

The association of Myc with the TIP60 complex was also shown to be mediated by E1A-TRRAP interaction [115,116], leading to the activation of target genes that boost ribosome biogenesis and non-coding RNA metabolic processes [116].
In conclusion, E1A-p400 binding could lead to extensive epigenetic changes in the host cell, as observed for e1a-Rb/p300 interactions.

Effects of Alu activation on gene expression

The biological role of *Alu* overexpression in response to virus infection is still an intriguing unsolved issue. *Alu* RNAs can act as modulator of gene expression, and phenotypic effects in response to exogenous *Alu* overexpression are starting to be elucidated.
Alu RNAs are known to exert their regulatory function both in the nucleus and in the cytoplasm. In the first case, in vitro studies showed that *Alu* RNAs bind Pol II and inhibit transcription during heat shock response [119]. Indeed, *Alu* RNA interferes with the association of Pol II with the promoter DNA, inhibiting the formation of the closed state of the preinitiation complex [119,120] (Figure 22A). Gene expression could also be influenced by *Alu* RNAs during mRNA maturation. For example, Pol III regions containing an *Alu* element were found highly homologous to the intron of a nearby coding gene. It was suggested that sequences transcribed from these Pol III regions could base pare with the corresponding homologous intronic sequence on the precursor mRNA, suggesting an antisense inhibition mechanism of mRNA maturation. However, the molecular mechanism of gene repression is not known in detail [121].
A very recent hypothesized mechanism of the effect of *Alu* elements on gene expression would rely on their potential influence of the spatial organization of the chromatin. Indeed, it is thought that SINE elements act as enhancers with *cis*- and

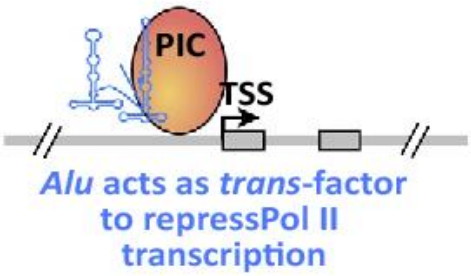
trans- mechanisms. The transcription of a SINE sequence that bears enhancer marks (Fos^{RSINE1}) was shown to increase the expression of its associated coding gene (Fos), possibly through interaction of the enhancer SINE eRNA with Pol II at the Fos promoter. This interaction would direct the Pol II machinery to sites of active transcription, hypothesizing a possible effect of the SINE eRNA to induce chromatin changes [122]. Consistent with the idea of an involvement of retroelements in the spatial organization of the chromatin, a very recent work suggests that *Alu*/B1 and L1 repeats might influence nuclear chromatin organization both with *cis* and *trans* mechanisms. Genomic L1 and *Alu*/B1 loci are subjected to homotypic clustering, being present in internal (transcriptionally active) and peripheral (transcriptionally inactive) nuclear compartments, respectively. In addition to DNA clustering, chromatin spatial organization would be influenced by *Alu*/B1 and L1 RNAs. Indeed, mouse L1 transcripts were shown to be able to bind their own DNA and the heterochromatin protein 1 (HP1α), suggesting a role of L1 transcripts in heterochromatin formation [123] (Figure 22B).

Lastly, a *cis*-role of repeated elements in influencing the organization of the nuclear space is thought to take place during serum starvation. TFIIIC was shown to re-localize to *Alu* elements in starvation conditions, inducing *Alu* H3K18 acetylation through TFIIIC intrinsic HAT activity. The new activated *Alu*s would be able to interact with CTCF sites nearby cell cycle genes through chromatin looping, ensuring an efficient proliferative response when serum is newly available [124].

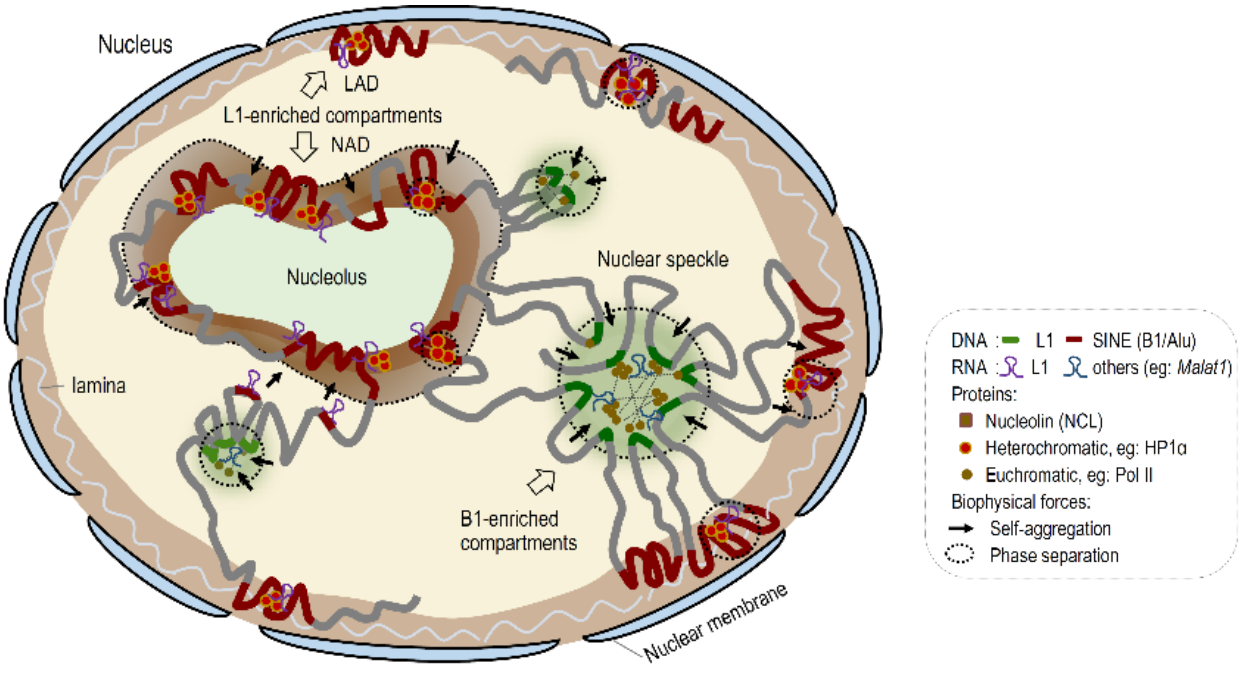
Modulation of gene expression by *Alu* transcripts can occur in the cytoplasm by regulating mRNA stability and/or the rate of translation. In the first case, the intermolecular interaction between two molecular species is thought to drive mRNA decay. An *Alu* element embedded in the 3′ UTR of a Pol II gene could base pair with an *Alu* element located either in the 3′ UTR of another coding transcript or in a long non-coding RNA (lncRNA). The imperfect intermolecular base pairing would lead to the formation of double stranded RNA, which can be recognized by the Staufen1

(STAU1) protein. The consequent recruitment of the ATP-dependent RNA helicase UPF1 triggers Staufen1 Mediated mRNA decay response, leading to RNA degradation [125,126] (Figure 22C).

Lastly, cytoplasmic *Alu* RNAs are known to enhance or inhibit translation initiation both *in vitro* and *in vivo* (Figure 22D). The right arm of free *Alu* RNAs was demonstrated to have a stimulatory effect on synthetic mRNAs [127,128]. Importantly, the enhancement of translation initiation was shown to be specific to a reporter gene and did not influence the global rate of protein synthesis, suggesting a mechanism that is independent than translational inhibition operated by the Protein kinase RNA-activated (PKR) [128].

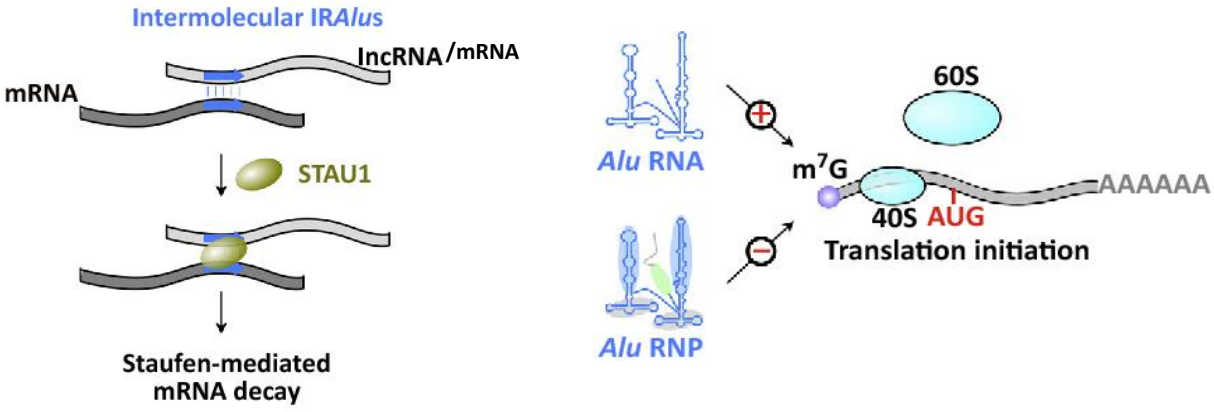


(A)



(B)

Figure 22. ***Continues on next page.***



(C) (D)

Figure 22. Possible mechanisms of gene regulation exerted by *Alu*/retroelement transcripts (*continues from previous page*). (A, from previous page) The right arm of *Alu* RNA is thought to have a repressive effect on Pol II transcription, inhibiting the formation of the preinitiation complex. (B, from previous page) L1 influences chromatin spatial organization with *cis* and *trans* mechanisms. L1 DNAs cluster together in euchromatic regions, whereas B1/*Alu* cluster in eterochromatic regions. Self-assembly of the DNA repetitive sequences would serve as anchor sites for transcription machineries, regulatory proteins and RNAs. The repetitive transcript would aggregate with RNA-binding proteins and with its own DNA, additionally stabilizing the structure of the nuclear compartment. Taken from [123]. (C) Two *Alu* sequences embedded in Pol II transcripts can base pair. STAU1 is then recruited at the dsRNA site, triggering Staufen1-mediated mRNA and consequently RNA degradation. (D) Effect of *Alu* RNA on mRNA translation. Free *Alu* RNAs and *Alu* RNP (Alu transcripts bound to the SRP particle) influence translation at the initiation step but with opposite effects: free *Alu* RNA enhance whereas *Alu* RNP inhibits translation initiation. All figures were adapted from [129]

On the other hand, *Alu* RNA bound to the SRP9/14 heterodimer (*Alu* SRP) has an inhibitory effect on protein synthesis [127,130] (Figure 22D). *Alu* SRP inhibits translation initiation by binding to the small ribosomal subunit (40S) and inhibiting its association with the mRNA [130], with the left arm showing a higher affinity to the SRP9/14 heterodimer than the right arm [80]. The opposite effects of *Alu* RNA and *Alu* SRP on translation could be explained with conformational changes of free *Alu* RNA or *Alu* RNA bound to the SRP heterodimer. Indeed, *Alu* RNA was demonstrated to exist in a relaxed folding state when not bound to the SRP, whereas a more compact folding structure is observed in the presence of SRP [131,132].

An effect on translation exerted by *Alu* RNA would also be mediated by *Alu* interactions with PKR, contrarily to what was reported by Rubin et al. [128]. Indeed, enhancement of protein synthesis was observed in response to *Alu* overexpression obtained through transient transfection or cell stress due to heat shock, Adenovirus infection or cycloheximide treatment. The physical interaction between PKR and *Alu* RNA, as detected from *in vitro* and *in vivo* experiments [133], would inhibit PKR autophosphorylation and its consequent activation, allowing protein synthesis [133]. This mechanism of action resembles viral mechanisms in counteracting host cell defence through the non-coding small RNA VAI of Adenovirus. PKR is present as a monomer in the inactive form, whereas the dimerization is stimulated by double stranded RNA generated from the transcription of both strands of the viral genome. The PKR dimer activates itself through autophosphorylation, subsequently inducing the phosphorylation of the α subunit of the Eukaryotic Initiation Factor 2 (eIF2α) and inhibiting host and viral protein synthesis. As a counter measure, VAI RNA, which is transcribed by Pol III and therefore not subjected to translation inhibition, binds PKR and inhibits its autophosphorylation. Therefore, VAI contributes to protein synthesis and viral replications (Figure 23).

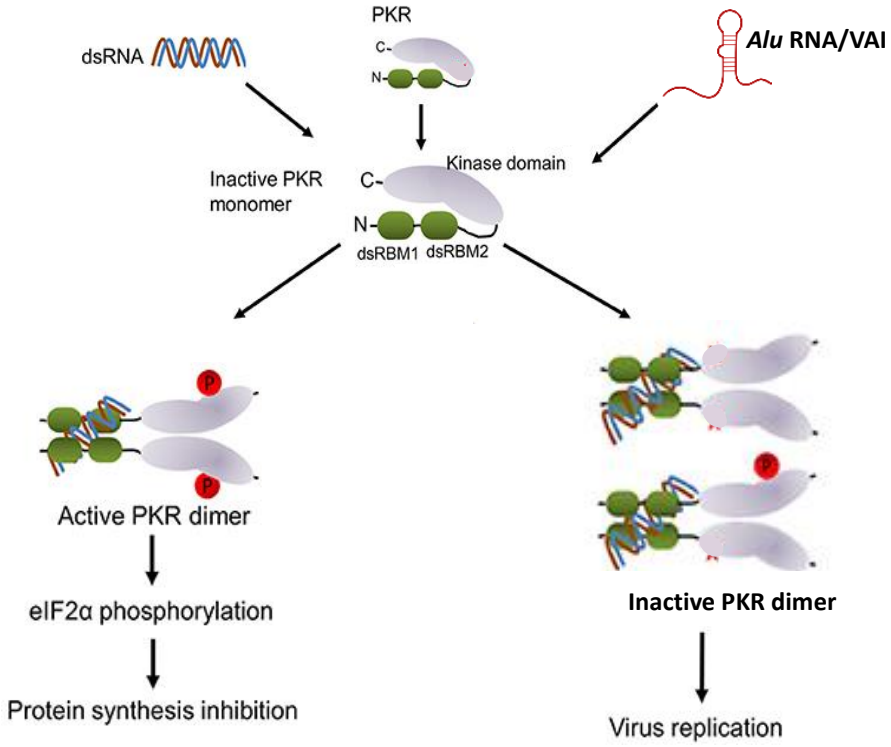


Figure 23. ***Legend on next page.***

Figure 23. Modulation of PKR activity by viral and *Alu* RNA *(figure from previous page)*. Viral dsRNA stimulates PKR dimerization and autophosphorylation, resulting in phosphorylation of eIF2α and subsequent inhibition of translation. *Alu* RNA and viral VAI RNA have a different effect on PKR, inhibiting autophosphorylation and therefore allowing protein synthesis. Taken from [134].

As seen previously, *Alu* RNA can modulate gene expression through different mechanisms. Therefore, it is reasonable to think that *Alu* transcripts might be involved in different cell functions and might influence cellular phenotypes.

Transcripts arising from *Alu Sc* elements were hypothesized to mediate oxidative stress typically observed in hyperglycemia, possibly through downregulation of genes involved in the oxidative stress response (eNOS and SOD2 genes). Interestingly, it was suggested that the phosphorylation of the transcription factor NF-κB was involved in *Alu*-mediated oxidative stress response, leading to an increased expression of Interleukin 1 beta (IL-1β). However, the molecular mechanism relying NF-κB phosphorylation was not investigated in detail [135].

A clearer view on the effect of the molecular interaction between NF-κB and Pol III transcripts is obtained from studies on murine SINEs. MHV68 infection of mouse fibroblasts enhances B2 SINE RNA, which is recognized by the RIG-I-like receptor leading to the activation of the mitochondrial antiviral-signalling protein (MAVS). MAVS would therefore bind the inhibitory subunit of NF-κB (IKKβ), resulting in the phosphorylation of the RelA subunit of the transcription factor NF-κB, inhibiting NF-κB transcriptional response. Additionally, the transcription activator RTA is also phosphorylated by IKKβ, leading to the activation of the viral genes [76,136,137] (Figure 24).

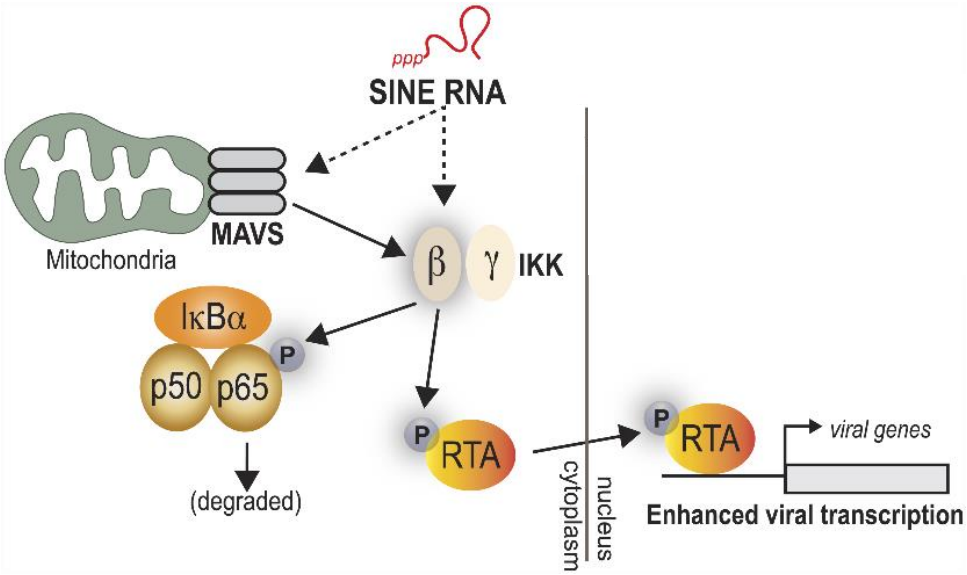


Figure 24. SINE RNAs activate the expression of viral genes in MHV68 infected cells. SINE RNA binds to the mitochondrial antiviral-signalling protein (MAVS), which activates IKKβ and the consequent phosphorylation of the p65 subunit of the NF-κB complex, consequently inhibiting NF-κB transcriptional activity. Simultaneously, IKKβ promotes RTA phosphorylation and the enhancement of viral gene transcription. Additionally, SINE RNA might stimulate IKKβ independently than MAVS. Adapted from [76].

Importantly, these findings suggest a functional relevance of virus-induced SINEs in enhancing virus gene transcription, through exploitation of the pathways involved in the cellular immune response to virus.

Alu RNA was also found implicated in the immune response mediated by the NLRP3 inflammasome. The overexpression of *Alu* RNA activates the MyD88 protein, triggering the NLRP3 inflammasome response and eventually leading to retinal pigment epithelium degeneration [138]. The supposed *Alu* RNA cytotoxic effect would be hindered by RNA degradation mediated by the DICER1 enzyme [139]. In addition to the epigenetic mechanisms described in the previous paragraphs, these findings support a tight control of *Alu*s expression also at the post-transcriptional level.

Lastly, *Alu* RNAs were demonstrated to be involved in the production of cytokines in systemic lupus erythematosus (SLE) syndrome. The Ro60 autoantigen was found associated with *Alu* RNA, and depletion of this protein led to an increased level of *Alu* transcripts and to the activation of cytokines, establishing a link among *Alu* elements, the autoantigen Ro60 and the immune response [140].

As seen above, *Alu* elements can regulate gene expression through a multitude of molecular mechanisms, resulting in a variety of downstream phenotypic effects.

For example, the transfection of an in vitro transcribed *Alu* RNA, induced epithelial-to-mesenchymal transition (EMT) in non-metastatic colon cancer cell lines (SW480 cells). Importantly, the metastatic counterpart cell line SW620 reported a reduction of the DICER1 enzyme and a consequent higher expression of *Alu* RNAs, supporting a role of *Alu* transcripts in cancer progression [141]. In contrast to this, several works report a decreased cell proliferation and the stimulation of cell differentiation in response to *Alu* RNA overexpression. Cell differentiation of pluripotent stem cells was induced through the transient transfection of an *Alu* sequence located upstream the NANOG gene, possibly through repression of the pluripotency genes OCT4 and NANOG [142]. *Alu* located proximally the NANOG gene are called DR2 *Alu* and are processed by DICER1 in small transcripts. Consistent with the previous study, upregulation of DR2 *Alu* mediated gene repression of the pluripotency NANOG gene, probably through base pairing of the processed *Alu* RNA to the 3' UTR of NANOG. This resulted in cell differentiation and a reduced cell proliferation of prostate cancer cells [143].

In line with the previous studies, the transfection and consequent overexpression of an *Alu*-like RNA induces differentiation of neuroblastoma cells, as suggested by a decrease in cell proliferation, increase in cell adhesiveness and a reduced malignancy [144].

Overall, *Alu* RNAs are involved in different pathways of the cellular immune system, whereas the phenotypic effects of *Alu* overexpression are not clear yet. Indeed, *Alu*

overexpression was reported to stimulate tumor progression toward a metastatic state, whereas several works support the involvement of *Alu* RNAs in differentiation and malignancy reduction of stem cells and tumor cells. An interesting clarification would result from studies on the proliferation of primary cells in response to *Alu* overexpression, even though no analyses have been performed in this direction to date.

Aim of the project

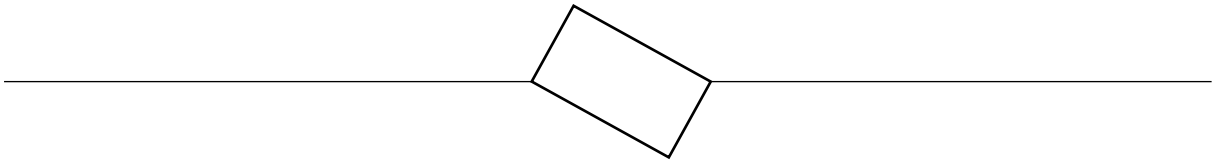
The presence of *Alu* retrotransposons in the human genome is a mystery that still has to be solved.

Alu elements are still actively retrotransposing in the human lineage. Indeed, the tight epigenetic silencing is thought to be aimed at the suppression of *Alu* insertional mutagenesis effect. However, the exceptional amplification efficiency through the human evolution and the retention in gene-rich regions suggest a positive role of *Alu* elements in cell processes, even though no clear function has been identified to date. Interestingly, *Alu* epigenetic silencing is released during different cell stresses such as heat shock, cancer progression and virus infection. Adenovirus 5 is one of the most studied viruses that are known to induce *Alu* overexpression. However, the virus factors that induce *Alu* activation are only partly understood, whereas transcriptional profiling of single *Alu* sequences that are overexpressed in response to virus infection has never been performed. An even more challenging issue is the clarification whether *Alu* overexpression represents a cell defence mechanism to virus infection, or instead if *Alu* transcripts are exploited by the virus to drive its own replication.

The first part of this PhD thesis is aimed at analysing the molecular mechanism of *Alu* derepression upon virus infection. e1a protein of Ad5 is known to induce a wide epigenetic reprogramming in human fibroblasts, therefore we reasoned that e1a might play a key role in *Alu* activation. e1a is known to induce epigenetic changes through interactions with the host proteins Rb and p300. The interaction with p400 was also demonstrated to be essential for cell transformation, even though the molecular mechanisms are not clear. We hypothesized that e1a might derepress *Alu* sequences through interactions with Rb, p300 or p400.

Alu expression was analysed in human fibroblasts infected with an Ad5 mutant that expresses only small e1a. *Alu* expression profiling was performed through RNA-seq

and by applying a bioinformatic pipeline that allows the detection of Pol III *Alu*s at single locus resolution. Secondly, *Alu* expression profiling was performed in fibroblasts infected with e1a mutants that are not capable to bind Rb, p300 or p400, with the final goal to identify host e1a interactors that are involved in *Alu* activation.
The second part of this work is aimed at identifying new biological roles of *Alu* elements, with the final goal to clarify the functional significance of *Alu* overexpression in response to virus infection or other stimuli. Two *Alu* sequences belonging to the *Alu S* subfamily were overexpressed through the generation of stable cell lines obtained through lentivirus infections. *Alu* overexpression was carried out in primary (human fibroblasts) and tumor (HeLa) cell lines, and the effects of *Alu* overexpression were assessed through gene expression analyses of RNA-seq data. The results were further confirmed by cell cycle analyses through flow cytometry. Finally, the potential mechanism of *Alu*-induced gene dysregulation was addressed through bioinformatic analyses on the 3' UTR of differentially expressed genes, and through the *in silico* prediction of miRNA targeting.



Chapter II

Alu RNA profiling in Adenovirus 5 infected cells

Results

Alu sequences are activated in response to different external stimuli, such as Adenovirus 5 infection, as already reported in the Introduction. However, it is still unclear which *Alu* loci are activated and to what extent, whereas the molecular mechanism that is responsible of this activation is still completely unknown. Since the viral early e1a protein plays a key role in inducing chromatin remodeling in human fibroblasts, we decided to investigate its precise role in activating *Alu* transcription. Moreover, the contribution of the e1a-interacting host proteins Rb, p300 and p400 in *Alu* activation in response to e1a stimulation is also analysed at single locus resolution.

The Adenovirus 5 dl1500 mutant expresses only small E1A

As already discussed in the Introduction, small E1A (e1a) protein is able to induce contact inhibited fibroblasts entry into the S-phase without activating the expression of any other protein of the viral genome. In order to dissect the activation of *Alu* sequences by Ad5, we decided to study the effects of e1a in *dl*1500 Ad5 mutants, evaluating the contribution of each e1a-interacting host protein in *Alu* activation.

The *dl*1500 mutant has a 9-bp deletion in the splice site of the unique E1A transcript (Figure 1), removing the last two bases of the 5′ exon of large E1A mRNA and the first seven bases at the 5′ end of large E1A transcript [92]. This results in the production of a virus mutant that is defective of replication in 293 and HeLa cells. The *dl*1500 mutant expresses e1a at a similar level as the wild type, whereas E1B is expressed at a considerably lower level respect to wt Ad5 [92]. e1a is capable to induce contact-inhibited human lung fibroblasts (IMR90 cells) entry into the S-phase 20hr post-infection, as assessed through thymidine incorporation assay [86].

The *dl*312 mutant has a 1030-bp deletion (Figure 1) (nucleotides 448-1349 of the viral genome) and produces virus particles that are defective of viral DNA synthesis and

E1A expression, therefore does not have any transforming activity [145,146]. This mutant will be considered as a negative control in the experiments described below.

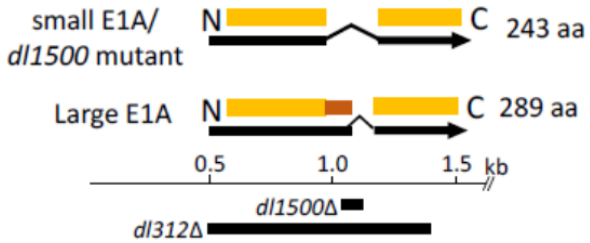


Figure 1. *dl*1500 and *dl*312 mutants of Adenovirus 5. The *dl*1500 mutant bears a 9-bp deletion, preventing a splicing event that leads to the production of the large E1A mRNA. Consequently, only the small E1A isoform is produced. A 1030-bp deletion abolishes the synthesis of both large and small E1A, resulting in no viral protein synthesis (*dl*312 mutant). Taken from [86].

Experimental and bioinformatic strategy developed to detect Alu transcripts

The analysis of *Alu* transcription is complicated by characteristic features of these elements. First of all, the repetitive nature challenges the unambiguous mapping of *Alu* RNA to the human genome; second, since *Alu* elements can be present inside Pol II-transcribed genes, *Alu* transcription could be the result of a Pol II activity of the *Alu*-hosting gene rather than a specific Pol III transcription; third, the low level of *Alu* expression requires a specific sequencing depth. A combination of specific RNA-sequencing (RNA-seq) features and the application of a specific bioinformatic pipeline previously developed in our laboratory, allows to uniquely map Pol III-transcribed *Alu* loci.

The application of a 100-nucleotide stranded paired-end sequencing facilitates the unique mapping of *Alu* reads. Indeed, RNA-seq fragments will be mapped to *Alu* loci based both on the nucleotide sequence and on the colocalization of the two 100-nt paired reads on the same locus. The 100-nt read length requirement is necessary to span a region that is most likely to encounter sequence variations within the repetitive *Alu* elements. Moreover, strand information lowers the probability to wrongly map fragments on different loci than what they originally come from.

The previously developed pipeline for the identification of transcriptionally active *Alu* sequences [147] allows to discriminate between a genuine Pol III *Alu* transcript and *Alu* RNAs putatively transcribed by Pol II. Indeed, *Alu* elements that present a similar read density on their body as well as their flanking regions are not considered in downstream analyses (Figure 2).

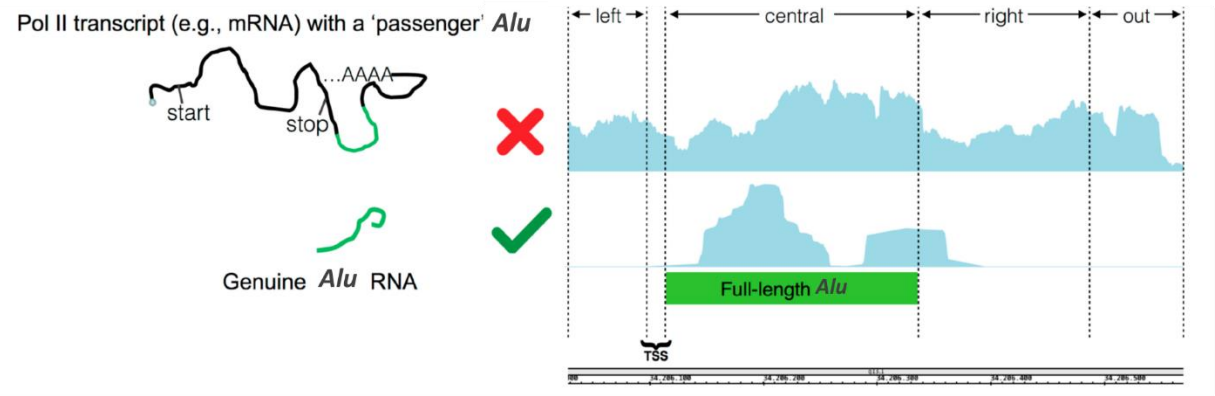


Figure 2. Application of the flanking region filter on *Alu* transcripts. *Alu* sequences embedded in mRNA show a coverage expression profile that extends upstream (left) and downstream (right) the *Alu* body. These *Alu* sequences will not pass the flanking region filter, therefore will not be taken into account for downstream analyses. Adapted from [147].

Moreover, in order to minimize the possibility to detect false positive Pol III transcribed *Alus*, only annotated sequences that are located in intergenic regions or within gene coding regions but in an antisense orientation were taken into account during reads alignment. The analyses presented in this work were performed considering only this subset of annotated *Alus*, which for brevity will be called "intergenic/antisense" *Alus*.

Lastly, given the low abundance of *Alu* transcripts, we performed RNA-seq analyses with a sequencing depth of 60 million paired reads per sample.

Small E1A stimulates Alu transcription

2-day contact-inhibited fibroblasts were either infected with Ad5 *dl*1500 or *dl*312 mutant at Multiplicity Of Infection (MOI) of 40, or mock-infected. Each infection was performed in duplicate. *Alu* profiling was assessed through RNA-sequencing and by applying the above described bioinformatic pipeline. Only *Alu* sequences that map at intergenic/antisense regions were taken into account, in order to minimize the probability to detect Pol II transcribed *Alus*. A total of 1880 *Alus* were detected as expressed (expression-positive) in the three samples, of which almost 57% were present only in *dl*1500 infected cells (which hereafter will be called "*dl*1500-specific *Alus*") and 37% were also expressed in *dl*312/mock cells (which hereafter will be called "common *Alus*"), whereas only 6.2% was exclusively detected in *dl*312/mock infected cells (Figure 3).

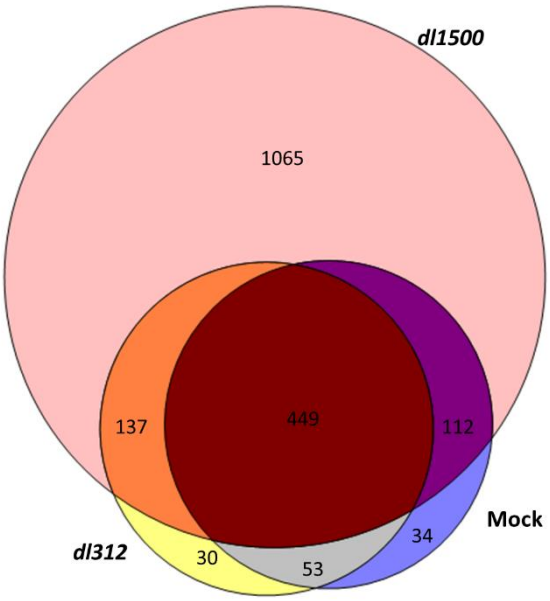


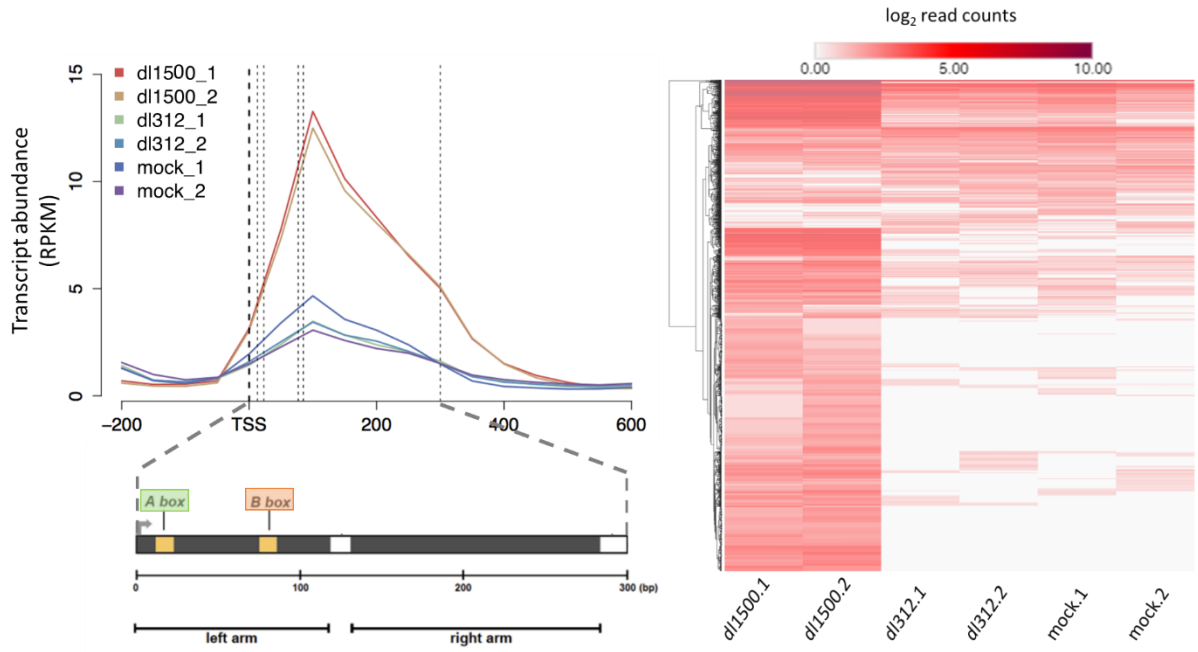
Figure 3. Number of expression-positive *Alu* sequences in *dl*1500, *dl*312 or mock-infected fibroblasts.

Alu induction/overexpression by e1a is a very limited phenomenon at the genome-wide scale, since only 0.22% of the intergenic/antisense *Alu* sequences could be detected as expression-positive by our experimental approach (Tab 1).

Table 1. Number of expression-positive *Alu* sequences in *dl*1500, *dl*312 and mock-infected cells. The percentage of expression of positive *Alus* was calculated compared to the total number of intergenic/antisense *Alu* sequences (799285 *Alu* loci).

	***dl*1500**	***dl*312**	**mock**
Number of expression-positive *Alus*	1763	669	648
Percentage of expression-positive *Alus*/annotated *Alus*	0.22%	0.08%	0.08%

Alu average expression was detected up to 4-fold increased by e1a infection (Figure 4A), and the *Alu* with the highest level of expression was detected with a maximum of 1100 normalized read counts (Figure 4B). As already shown in Figure 3, the heat map confirms that only a subset of *Alu* sequences is already transcribed in *dl*312 or mock samples, whereas the majority of *Alu* sequences is activated in *dl*1500 infected fibroblasts.



(A) (B)

Figure 4. e1a stimulates the transcription of *Alu* sequences. (A) Average expression profile of *Alu* sequences mapped against the annotated *Alu* element. The average read count abundance was normalized as Read Count Per Million Kilobase (RPKM) and analysed in *dl*1500, *dl*312 and mock-infected cells. Each infection was performed in duplicate. (B) Heat Map showing the transcript abundance of each *Alu* sequence detected as expression-positive in both replicates belonging to the same sample. Read Counts were normalized using the DESeq2 package in R.

We next considered if those *Alus* that are already expressed in uninfected cells (common *Alus*) are more prone to be activated by e1a stimulation than *Alus* that are detected as expression-positive only in *dl*1500-infected cells (*dl*1500 specific *Alus*). Interestingly, it seems that common *Alus* have a higher expression after e1a infection than *Alus* that are expressed only in *dl*1500 cells (Figure 5).

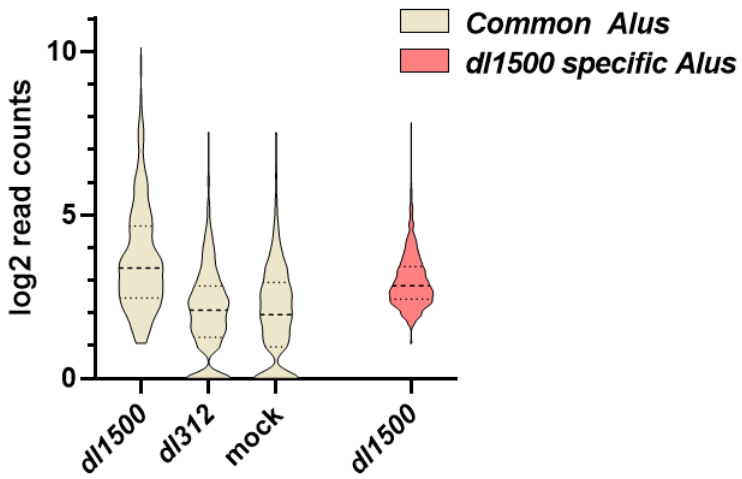


Figure 5. e1a efficiently stimulates the expression of *Alu* sequences that are already transcribed at a basal level in unperturbed cells. The violin plot shows the distribution of normalized read counts (as obtained from DESeq2) of common-*Alus* and *Alus* exclusively express in *dl*1500-infected cells. The solid line is the median and the dashed lines mark the Inter Quartile Range (IQR).

Features of Alu subfamilies

We next investigated the contribution of *Alu* sequences belonging to the oldest subfamily J, to the intermediate subfamily S, and to the youngest subfamily Y to the total expression of *Alu* sequences, in terms of number of expression-positive *Alus*.

We calculated the relative enrichment between the *Alu* sequences detected as expressed in *dl*1500, *dl*312 and mock and the number of annotated *Alus* belonging to each subfamily (Table 2). A significant enrichment in the number of expressed *Alu S* sequences was revealed compared to the annotated sequences, and the effect is even slightly increased in *dl*1500-infected cells. However, a significant decrease of enrichment in *Alu J* sequences compared to the pool of annotated intergenic/antisense *Alus* was observed in *dl*1500, consistent with the notion that the accumulation of mutations in the *Alu J* subfamily led to a transcriptional inactivation of these sequences [24].

Table 2. Number of *Alu* sequences belonging to the subfamilies *Alu J*, *Alu S* or *Alu Y* that are detected as expression-positive in *dl*1500, *dl*312 or mock samples. The abundance of *Alu* transcripts belonging to each subfamily was calculated against the number of annotated *Alus* located in intergenic or antisense regions. The *p*-value (2-tailed Fisher-exact test) was significant (*p*-value < 0.01) only for *Alu J* and *Alu S* subfamilies expressed in *dl*1500 infected cells.

	Alu J	***Alu S***	***Alu Y***	**Tot**
N. of annotated *Alus*	213441 (26.7%)	484941 (60.67%)	100903 (12.62%)	799285
N. of expressed *Alus* in *dl*1500	162 (9.18%)	1494 (84.74%)	107 (6.07%)	1763
N. of expressed *Alus* in *dl*312	123 (18.39%)	488 (72.94%)	58 (8.67%)	669
N. of expressed *Alus* in mock	122 (18.83%)	470 (72.53%)	56 (8.64%)	648
***Alu* enrichment in *dl*1500**	0.34	1.4	0.48	
(p-vAlue)	(2.5×10^{-3})	(2.1×10^{-4})		
***Alu* enrichment in *dl*312**	0.69	1.2	0.69	
***Alu* enrichment in mock**	0.7	1.2	0.68	

Alu expression is more enhanced in Alu S subfamily

Among the 1846 expression-positive *Alus* in either *dl*1500 or *dl*312 infected cells, a total of 424 *Alus* were found differentially expressed (DE) in *dl*1500 compared to *dl*312-infected cells, meaning that almost 23% of the detected *Alus* were significantly (*adj p*-value < 0.05) overexpressed in response to e1a induction. Even though the transcription of a minor percentage of *Alus* is significantly enhanced by e1a, *Alus*' overexpression seemed to be particularly efficient, reaching a 73-fold induction in some *Alu* sequences. DE *Alus* with the higher Fold Change belong to the *Alu S* subfamily, whereas the only two downregulated *Alus* belong to the J subfamily (Figure 6A). Interestingly, *dl*1500 specific *Alus* have a lower average Fold Change compared to common *Alus,* suggesting once more that those sequences that are already transcribed at a basal level in *dl*312 cells are more efficiently overexpressed than *Alus* that are not transcribed in unperturbed cells (Figure 6B). Even though the lower Fold

Change might result from a normalization artefact in DESeq2 calculations, this observation supports the existence of a pre-marked state of *Alu* sequences in uninfected cells. This pre-marking feature would facilitate *Alu* transcription in response to e1a stimulation.

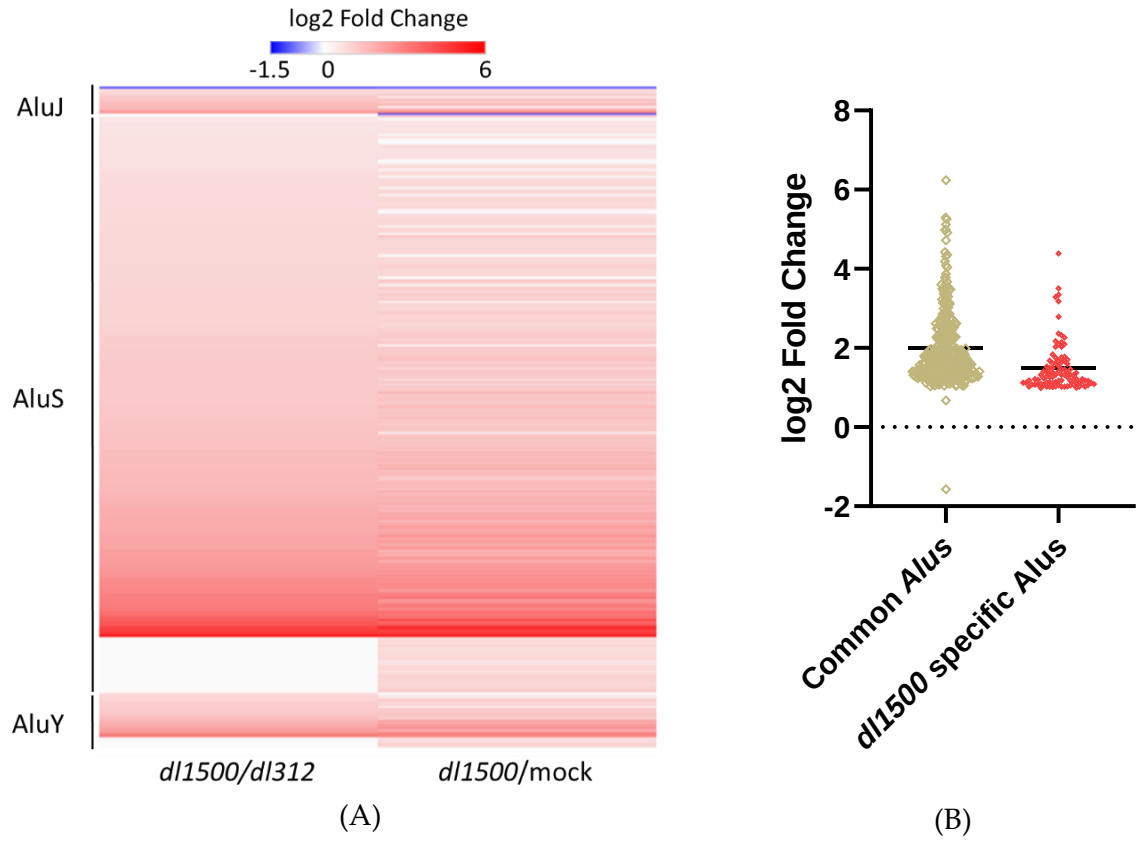


Figure 6. Overexpression of *Alu* sequences induced by e1a compared to *dl*312 and mock-infected cells. (A) Heat Map showing Differentially Expressed *Alu* sequences in *dl*1500 cells compared to *dl*312 or mock samples, as revealed from DESeq2 analyses. *Alu* families are sorted based on *Alu J*, *Alu S* or *Alu Y* subfamily. (B) Fold increase of *dl*1500/*dl*312 expressed *Alus*. *Alu* sequences were divided in common *Alus* (expression positive in *dl*1500 and *dl*312 cells) and *dl*1500 specific *Alus* (expression positive in *dl*1500 cells but not in *dl*312). The solid line marks the average $\log_2$ Fold Change value.

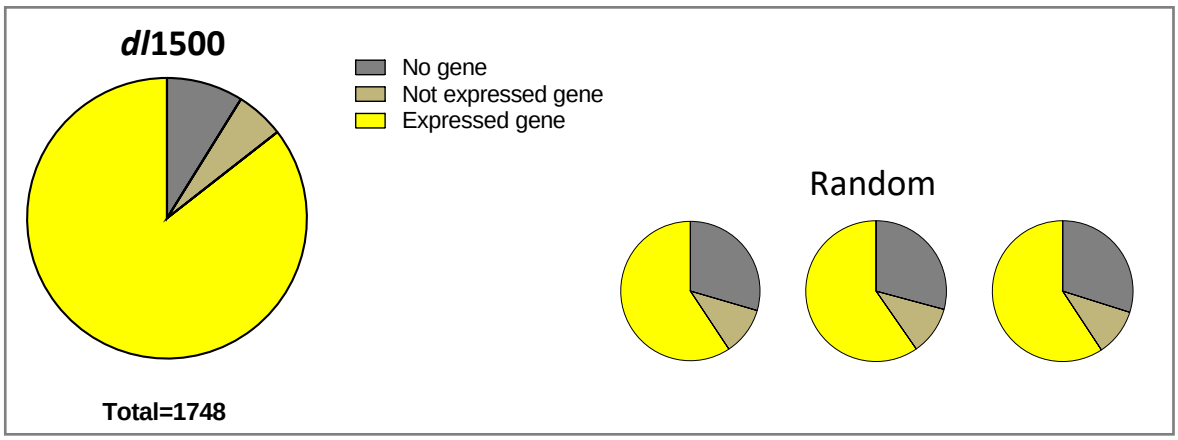
Genomic association of Alu sequences with coding genes

We next analysed the transcription of genes that are localized in the proximity of *Alu* sequences that are expressed in both *dl*1500 and *dl*312-infected cells.

Therefore, using the GREAT online software (http://great.stanford.edu) we analysed the association of expression-positive *Alus* with expressed genes in our dataset.

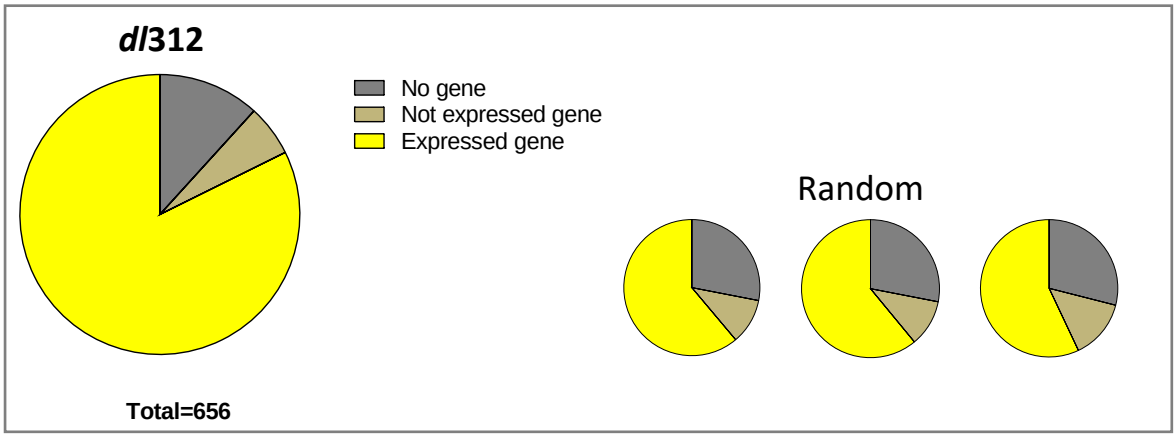
We defined a regulatory region associated to any annotated gene as 100-kb upstream and 100-kb downstream its TSS, assigning a maximum of two genes for each *Alu* sequence. Moreover, regulatory regions that fall outside the 100-kb rule were also taken into account if supported by experimental evidence, as implemented by the "curated regulatory domains" option in GREAT.

A total of 1748 expression-positive *Alus* in *dl*1500 and 656 expression-positive *Alus* in *dl*312 were intersected with GREAT regulatory regions, and the association of *Alu* sequences with no genes, not expressed genes and expressed genes was analysed. In both *dl*1500 and *dl*312 samples, the fraction of expression-positive *Alu* sequences associated with expressed genes is remarkably higher compared to random control sets of not expressed *Alu* sequences (Figure 7A and B).



(A)

Figure 7. ***Figure continues on next page.***



(B)

Figure 7. Genomic association of expression-positive *Alu* sequences with expressed genes *(figure continues from previous page)*. The number of *Alu* sequences associated with expressed genes increases compared to the same number of random (not expressed) *Alu* sequences. (A-B) GREAT analyses in *dl*1500 and *dl*312 samples. The same analyses were performed three times on the same number of not expressed *Alu* sequences belonging to the pool of intergenic/antisense annotated *Alus*. No gene=*Alu* not associated with any gene. Not expressed gene=*Alus* detected as associated with at least one gene from GREAT analyses, but this gene is not expressed in our dataset. Expressed gene=*Alu* associated with at least one expressed gene.

As shown in Figure 8, the increase of *Alu* sequences associated with expressed genes in both *dl*1500 and *dl*312 cells was statistically significant compared to a random control (Figure 8).

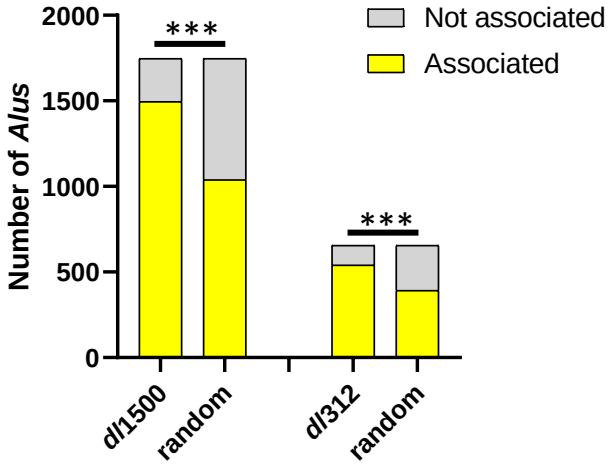


Figure 8. Expression-positive *Alu* sequences are located nearby transcribed genes. The number of expressed *Alu* sequences that are associated with expressed genes increased both in *dl*1500 and *dl*312 samples compared to a random control of not expressed *Alus*. A *Chi-squared* test was used to evaluate a statistical difference between *dl*1500/*dl*312 and a random control (which was calculated as the average of the three random controls shown in Figure 7). *** = *p*-value < 0.001

Overall, the data suggest that *Alu* expression is influenced by the transcription of genes that are located within 100-kb upstream or downstream the *Alu* central body, and this effect is independent of e1a activity.

e1a mutants defective in binding host proteins Rb, p300 or p400

It is well known that e1a induces a global epigenetic reprogramming of human contact-inhibited fibroblasts interacting with the host proteins Rb and p300, whereas the contribution of e1a-p400 interaction is not clear yet (See Introduction). We reasoned that e1a-induced epigenetic modifications could also influence *Alu*

expression in *dl*1500-infected cells, therefore we studied *Alu* activation in cells infected with e1a mutants that are not capable to bind Rb, p300 or p400.

The e1a-Rb binding mutant (called hereafter e1aRb$^-$) was obtained through amino acid modifications in the CR1 region and the LXCXE motif inside the CR2 region. Indeed, the amino acid substitutions L43A, L46A and Y47A, and the deletion of amino acids 112-128 are sufficient to inhibit e1a-Rb interaction [86].

e1a binding to p300 was impaired by the amino acid substitutions R2G, E59A, V62A, F66A and E68A, which eliminate the e1a-p300 interaction with the domain TAZ2 [86]. Lastly, amino acids 26-35 were substituted with alanines in order to eliminate e1a-p400 interaction [105,148] (Figure 10). e1a proteins that are not capable to bind p300 and p400 will be called e1ap300$^-$ and e1ap400$^-$, respectively.

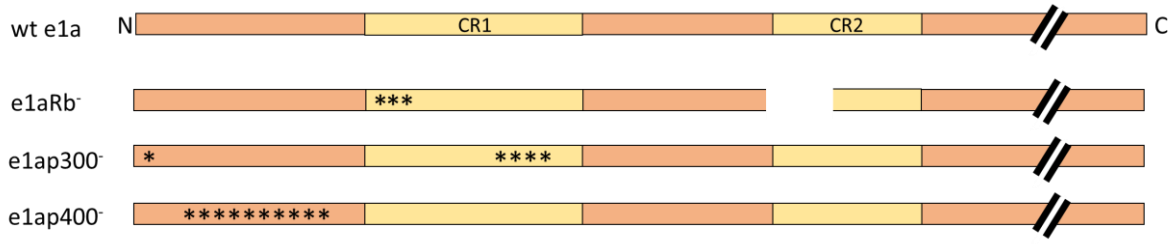


Figure 10. Schematic representation of e1a mutants that are not capable to bind Rb, p300 or p400. e1aRb$^-$ was obtained through amino acid substitutions in the CR1 region and amino acid deletions in the CR2 region, whereas e1ap300$^-$ and e1ap400$^-$ were obtained through amino acid substitutions in the N-terminal (e1ap400$^-$) and CR1 region (e1ap300$^-$).

Importantly, the substitution/deletion of e1a amino acids to obtain the three mutants did not interfere with the ability of e1a to bind the other two proteins, as verified by co-immunoprecipitation experiments [86,105,113,149].

Validation of the expression of small E1A in Ad5 binding mutants by western blot

We first determined the Multiplicity Of Infection (MOI) that is needed by e1a binding mutants to obtain the same levels of e1a protein as the wild type (MOI of 40 in *dl*1500 mutants, as performed for the RNA-seq experiment previously described).

Different MOIs were first tested in HeLa cells, spanning a range between 40-320 MOIs for e1aRb$^-$, 10-160 MOIs for e1ap300$^-$ and 10-60 MOIs for e1ap400$^-$. The levels of e1a proteins were measured by western blot using the M73 antibody, which recognizes e1a C-terminus end. Since all the mutants have aminoacid substitutions/deletions in the first half of the N-terminus region, the ability of M73 antibody to recognize e1a is not impaired in the three mutants. The Ku86 protein was used as loading control. As shown in Figure 11, the MOIs needed to reach the same protein level as obtained in *dl*1500 40 MOIs were selected as 160 for e1aRb$^-$, between 40 and 80 for e1ap300$^-$, and less than 10 MOI for e1ap400$^-$. We considered 60 MOIs and 6 MOI for e1ap300$^-$ and e1ap400$^-$, respectively.

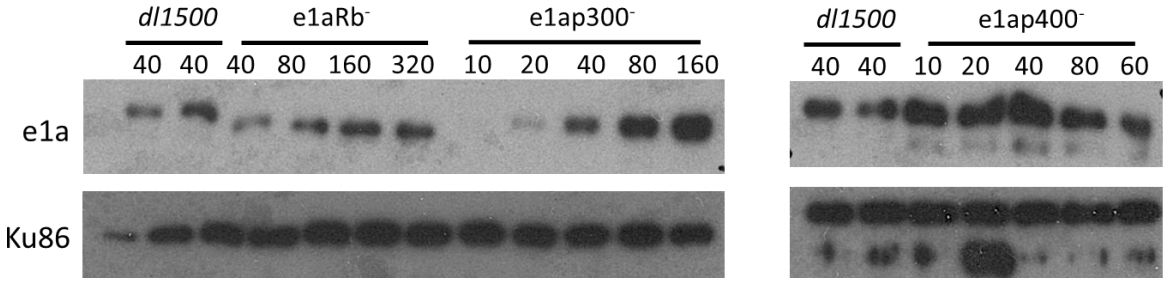


Figure 11. **Western blot showing the protein levels of e1a binding mutants obtained from virus infections of HeLa cells at different MOIs.** The western blot was performed in order to select the required MOI to obtain the same protein level as obtained for wt e1a (*dl*1500). The proper MOIs were selected as 160 for e1aRb$^-$, 60 for e1ap300$^-$ and 6 for e1ap400$^-$. Ku86 was used as a loading control.

Next, we tested whether IMR90 cells infected with e1a mutants at the determined MOIs in HeLa cells produced a comparable amount of e1a protein.

2-day contact-inhibited IMR90 cells were infected in duplicate with *dl*1500, e1aRb$^-$, e1ap300$^-$ and e1ap400$^-$ at an MOI of 40, 160, 60 and 6, respectively. After 24 hours, the cells were scraped off the plate and the protein lysate was analysed through western blot as for HeLa cells. e1a binding mutants expressed almost the same level of e1a as observed in *dl*1500 (Figure 12). The lower amount of e1aRb$^-$ protein is probably due to the inability of e1a to release the transcription factor E2F from Rb inhibition, therefore preventing e1a transcription to be increased by E2F. Indeed, e1aRb$^-$ could be efficiently synthesized in HeLa cells for the presence of the oncoviral E7, which is transcribed from the HPV18 virus DNA integrated into HeLa cells genome [150]. E7 binds Rb resulting in its degradation and functional inactivation [151], therefore E2F is constitutively active in HeLa cells.

The lower levels of e1aRb$^-$ in IMR90 cells are not likely to constitute a bias in downstream analyses, since e1aRb$^-$ mutants were shown to repress host genes even more than wild type e1a [86].

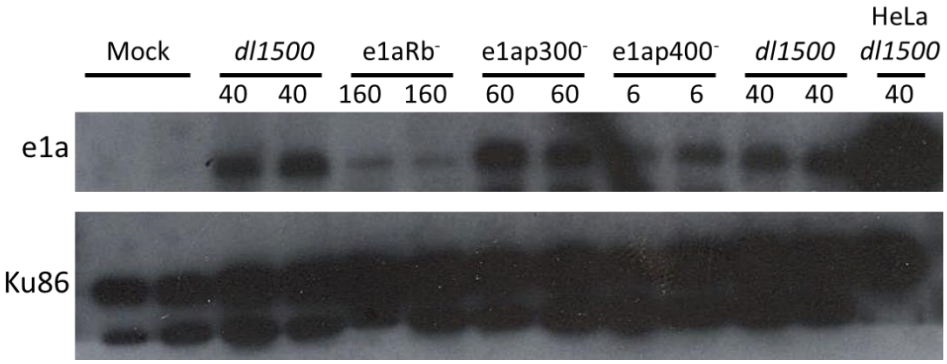


Figure 12. **Western blot showing protein levels of e1a binding mutants obtained from virus infections of IMR90 cells at different MOIs.** Legend as in Figure 11.

Modulation of Alu transcriptome by e1a binding mutants

In order to identify possible molecular targets of e1a that are involved in *Alu* activation, *Alu* expression profile was analysed through RNA-sequencing (as previously described) in fibroblasts infected with e1aRb$^-$, e1ap300$^-$ and e1ap400$^-$, as well as with the *dl*1500 mutant and mock-infected. *Alu* expression was analysed in three rounds of infections (considered as 3 biological replicates). *Alu* sequences are less activated when e1a is not capable to bind p400, whereas in the absence of Rb and p300 binding *Alu* sequences are more- or less- activated compared to wt e1a, depending on the replicate considered (Figure 13).

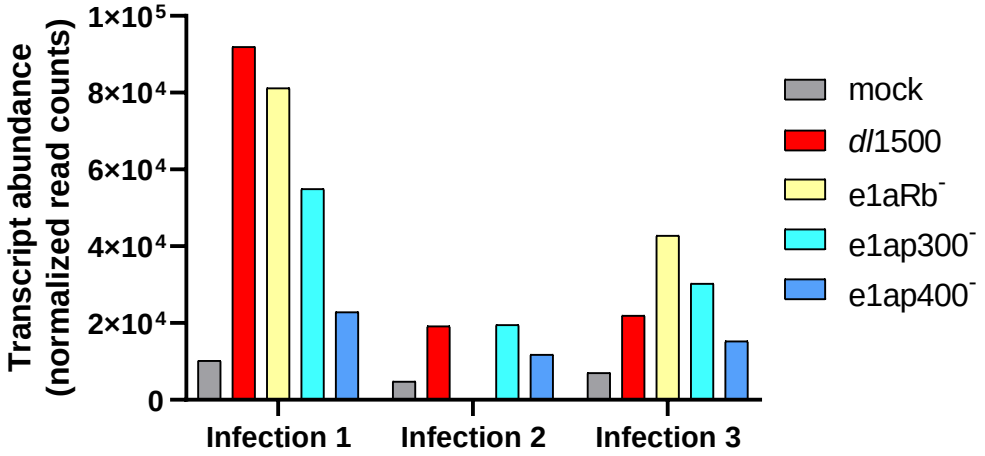


Figure 13. **Total expression of *Alu* sequences in *dl*1500, e1aRb$^-$, e1ap300$^-$ and e1ap400$^-$**. The total *Alu* expression was obtained summing the normalized read counts of each *Alu* sequence (only for *Alus* detected in least 2 replicates) in each sample. Normalized read counts were obtained through the DESeq2 package. Replicate 2 of e1aRb- mutant was missing due to failure in library construction.

Indeed, as shown by the heat map in Figure 14, e1ap400$^-$ *Alu* sequences are less overexpressed, or even downregulated, than in wt e1a, e1aRb$^-$ or e1ap300$^-$ (Figure 14).

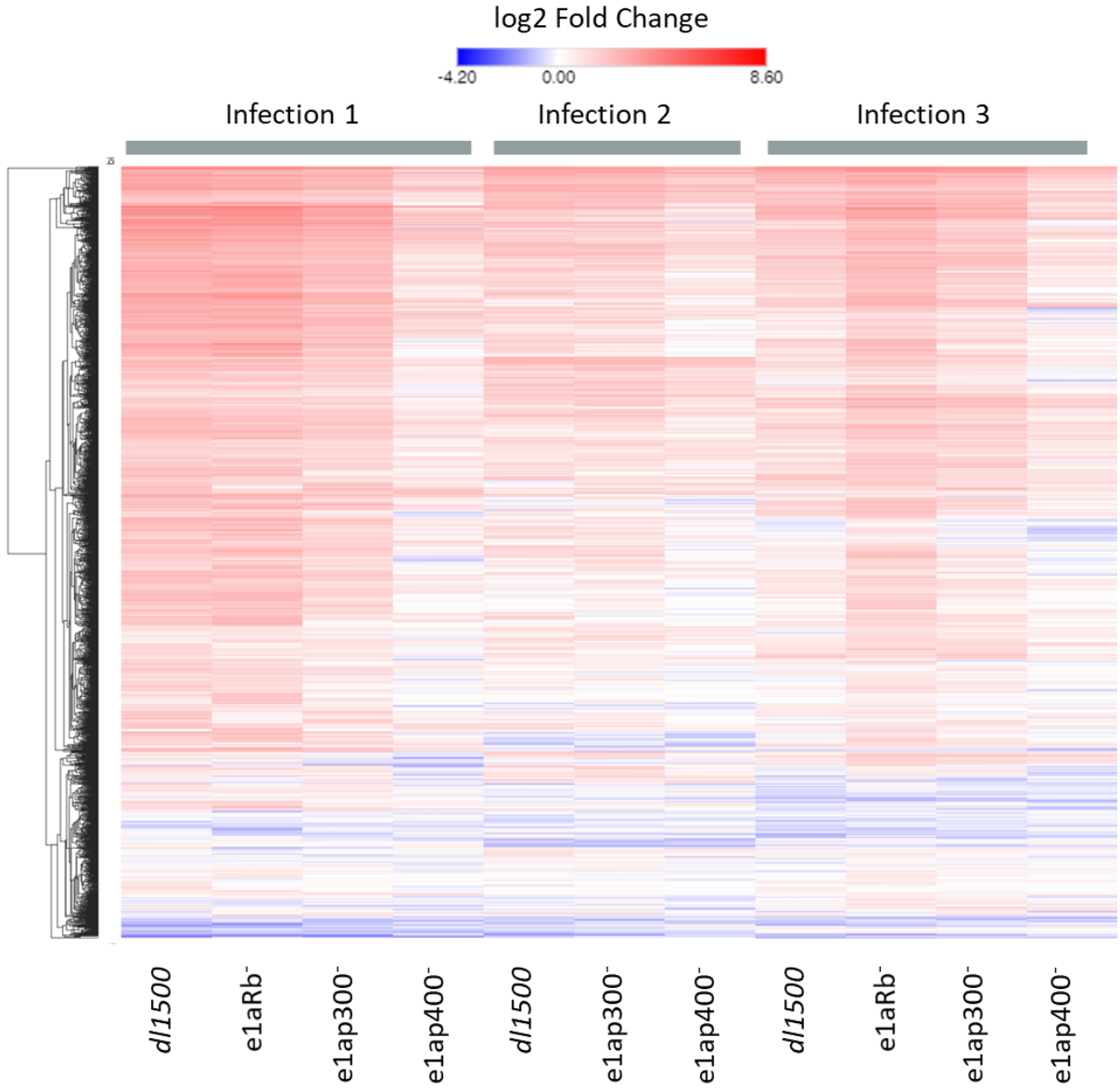


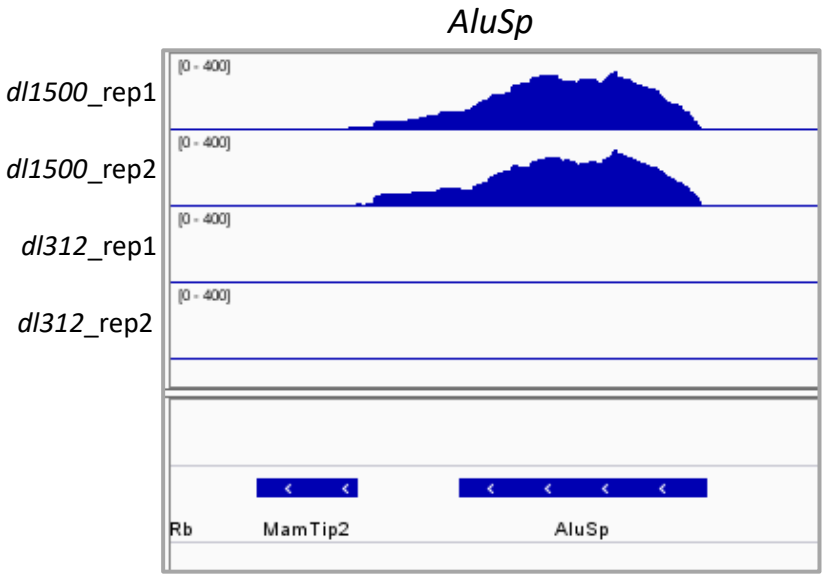
Figure 14. Heat Map showing the up- and down- regulation of *Alu* sequences compared to the corresponding mock infection controls.

Analysis of Alu activation in e1a mutants by Real Time PCR

In order to validate RNA-seq data, the expression of two *Alu* sequences was analysed through RT-qPCR. Two *Alu* sequences were selected as particularly suitable for this study. First, as detected by RNA-seq data, the sequences *AluSp* on chromosome 5 and *AluSc* on chromosome 18 are among the first 10 most overexpressed sequences

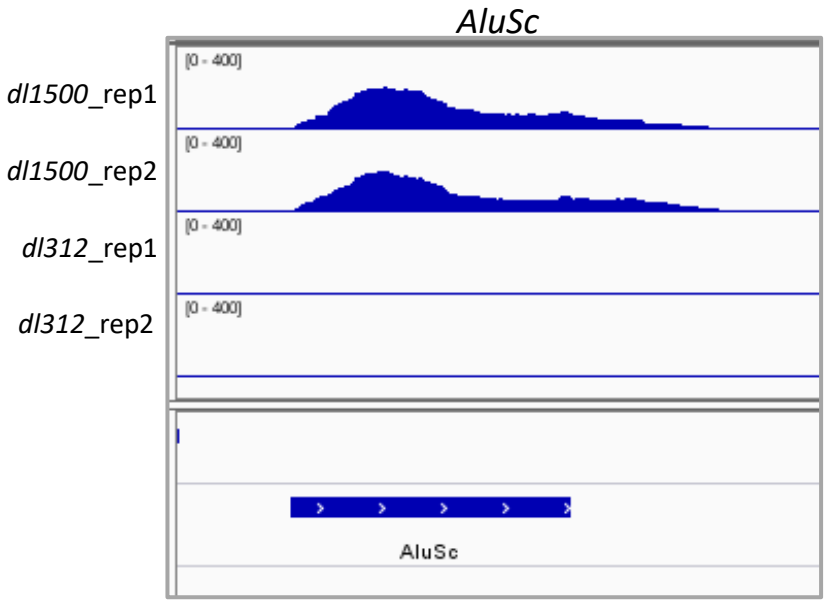
compared to *dl*312-infected fibroblasts. Moreover, from visualization in the genome browser it is possible to verify that the two selected *Alus* have a clear expression profile and a long 3' trailer (Figure 15). The existence of a relatively long stretch of nucleotides (80-100 nt) downstream the poly(A) tail is a very important feature for *Alu* expression analyses through RT-qPCR.

Since *Alu* elements are highly repetitive, the annealing of amplification primers on their body would not allow to distinguish between *Alu* transcripts originated from different genomic loci. Therefore, it is necessary to design primers that anneal on the unique *Alu* region that was incorporated after the retrotransposition process, indeed, the 3' trailer. The sequence and the genomic regions of the two *Alu* sequences of interest are available in Table S1 of Supplementary Information I.



(A)

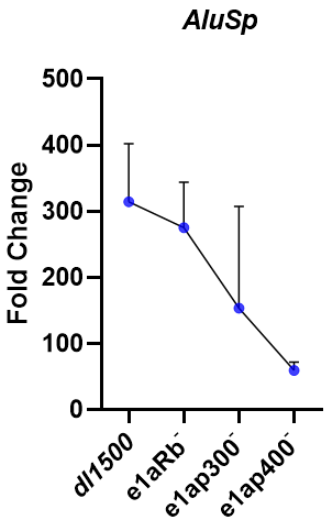
Figure 15. ***Continues on next page.***



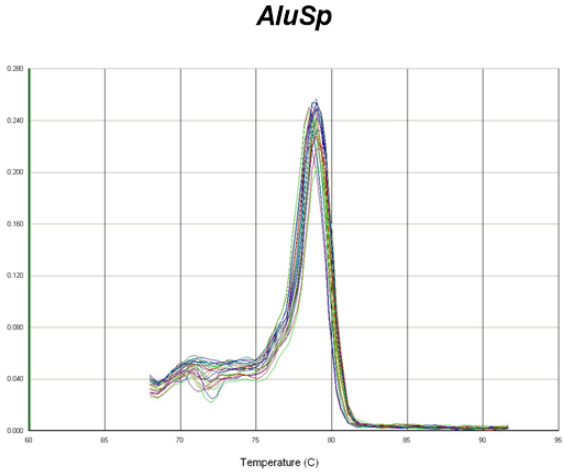
(B)

Figure 15. Visualization of *Alu* expression profile through the Integrative Genomic Viewer (IGV) software *(figure continues from previous page)*. The coverage expression profile of *AluSp* (A) and *AluSc* (B) is characterized by the presence of a long 3' trailer that extends beyond the annotated *Alu* element (squared box in the bottom part of A and B).

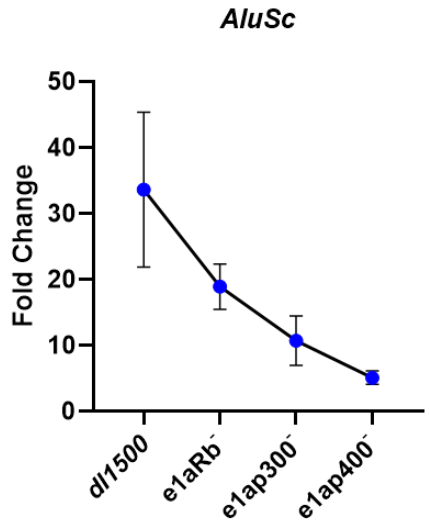
RT-qPCR data show a general trend towards a lower expression of the two tested *Alus* when cells were infected with e1a binding mutants, confirming that e1ap400^{-} is the least efficient mutant in activating *Alu* expression (Figures 16A and 16C).



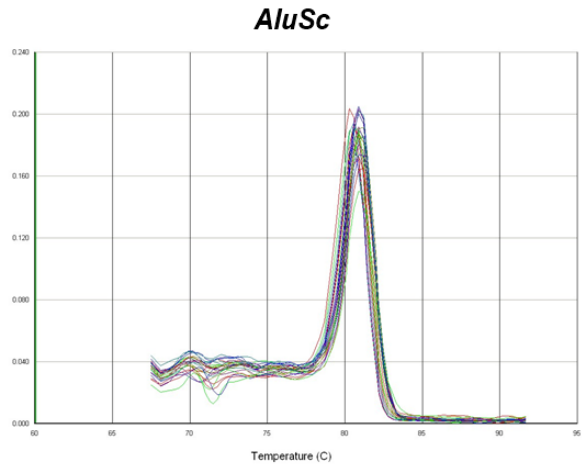
(A)



(B)



(C)



(D)

Figure 16. **Expression analyses of *AluSp* and *AluSc* through RT-qPCR in e1a binding mutants.** (A and C) *AluSp* and *AluSc* expression is enhanced in wt e1a-infected cells compared to mock-infections, whereas a lower expression was observed in e1a binding mutants. The U1 snRNA gene was used as internal standard. (B and D) The melting curves show the absence of aspecific amplification.

Analysis of Polymerase III genes activation by e1ap400-

From RNA-seq and RT-qPCR data, it seems that p400 is a major e1a interactor in enhancing *Alu* expression, possibly through e1a-induced recruitment of p400 on *Alu* elements. In order to investigate if this mechanism is also detectable for Pol III genes or is specific for *Alu* elements, we analysed the expression of the 7SL and U6 genes in *dl*1500 and e1ap400- infected cells through RNA-seq and RT-qPCR. Three 7SL gene variants exist in the human genome (RNA7SL1, RNA7SL2 and RNA7SL3), therefore a pair of primers was designed on the conserved regions (Figure 17).

Figure 17. **Alignment of the three 7SL gene variants in the human genome.** Red boxes indicated the annealing region of the primers used in RT-qPCR.

The expression of the Pol III genes 7SL and U6 did not change substantially in response to wt e1a stimulation or e1ap400- overexpression, as revealed by RNA-seq and RT-qPCR data (Figure 18).

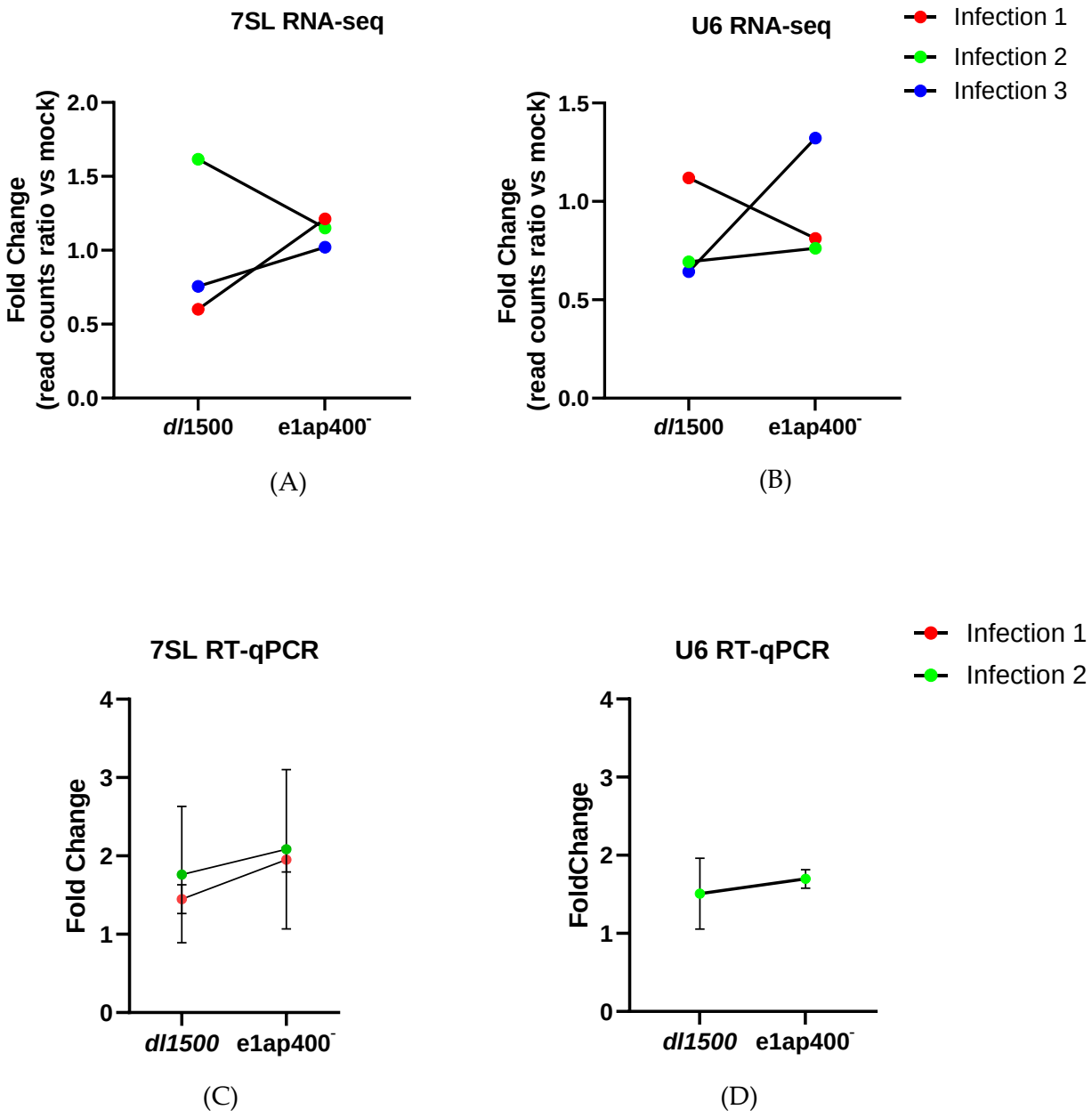


Figure 18. The expression of Pol III genes is not influenced by wt e1a or e1ap400-. The transcript abundance of the 7SL and U6 genes was compared to mock infected cells and detected through RNA-seq (A,B) and RT-qPCR (C,D). The U1 snRNA gene was used as internal standard in RT-qPCR analyses (C,D).

Overall, it seems that e1a-induction of *Alu* expression is due to molecular events specifically taking place at *Alu* sequences rather than to a general involvement of the Pol III transcription machinery.

Epigenetic context of expressed Alu sequences

The data reported above support a specific stimulation of *Alu* transcription, which most likely occurs through e1a-p400 binding. However, it is not known how e1a can specifically target *Alu* sequences instead of inducing a general transcription of Pol III genes. One explanation could rely on the existence of chromatin modifications that mark specific *Alus* to be targeted by e1a and to induce an efficient transcription of these pre-marked elements. Therefore, we analysed the enrichment of histone modifications and of histone acetyl transferases on *Alus* expressed in mock and *dl*1500 infected cells. ChIP-seq data on H3K18ac, H3K27ac, p300 and H3K4me1 were retrieved from Ferrari et al. [86] and Ferrari et al. [88] and peaks of enrichment were evaluated on the average sequence of *Alus* expressed in our dataset.

No significant enrichment of H3K18ac, H3K27ac and p300 was detected at *Alu* sequences expressed in mock and *dl*1500 infected cells. On the other hand, H3K4me1 was found slightly enriched at *Alu* sequences expressed in *dl*1500 as well as mock-infected cells, compared to an equal number of random *Alu* sequences (Figure 16). Therefore, H3K4me1 is likely to be an important feature of expressed *Alu* sequences: the enrichment in mock-infected cells suggests that this epigenetic mark is deposited independently of e1a, and H3K4me1 could mediate e1a recruitment to efficiently enhance *Alu* transcription. H3K4me1 is an active enhancer histone mark and is known to recruit Transcription Factors and chromatin remodeling complexes such as the TIP60/p400 complex [110]. Therefore, it is tempting to speculate that the combined presence of H3K4me1 and e1a could facilitate the recruitment of the TIP60/p400 chromatin remodeling complexes on pre-marked *Alu* sequences, ultimately leading to their activation.

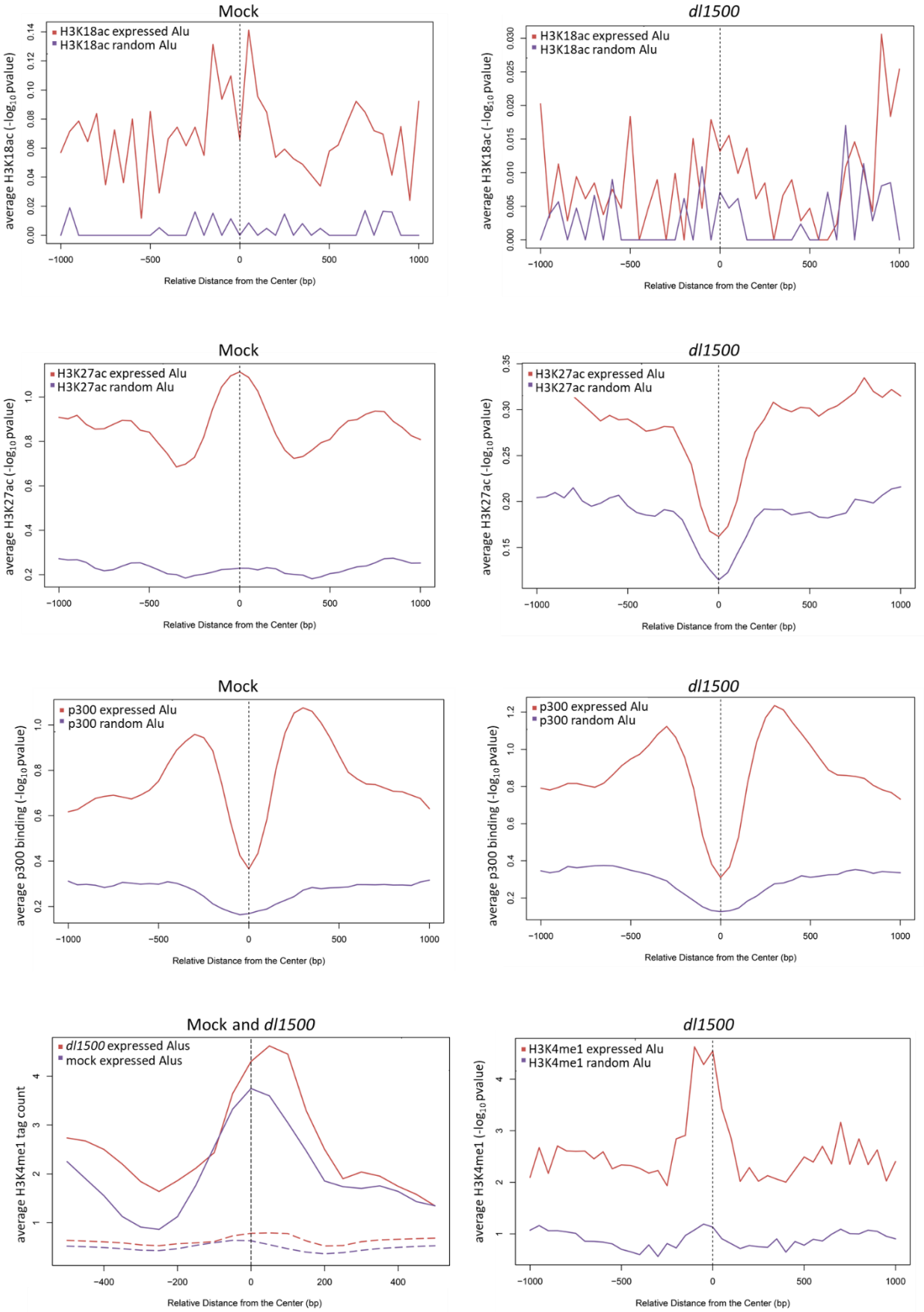


Figure 16. Enrichment of H3K4me1, but not of H3K18ac, H3K27ac and p300 on expressed *Alus*. The enrichment of H3K4me1 is shown both as tag count (mock and *dl*1500) and *p*-value (*dl*1500). The enrichment levels were compared against an equal number of random *Alu* sequences. Dashed lines in the bottom left panel indicate H3K4me1 average signal on all intergenic/antisense genomic *Alus*.

Materials and Methods

Construction of E1A-substituted Ad5 vectors

The *dl*1500 and *dl*312 mutants used in the first RNA-seq experiment were constructed as previously described [92,145]. The *dl*1500 and e1a constructs used in e1a binding mutants experiments were created through PCR mutagenesis and afterwards inserted into the shuttle plasmid pAdlox. loxP recombination between the Adenovirus 5 backbone Ψ5 and the shuttle plasmid pAdlox was performed as described in Hardy et al. [152].

Cell lines and virus infections

IMR90 cells and HeLa cells (ATCC) were grown in a humified atmosphere 5% CO_2 at 37°C in DMEM medium supplemented with 10% fetal bovine serum, 100 U/ml penicillin and 100 μg/ml streptomycin. IMR90 cells were grown in 60 mm plates until 90-100% confluency and then incubated two more days without changing the medium. On the day of infection, the medium was aspirated (conditioned medium) and the cells were incubated in 100 μl of virus diluted in PBS. The incubation was performed for one hour at 37°C, after which the cells were washed and the conditioned medium was added back for 24 hours (the use of conditioned medium instead of fresh medium is required to avoid a serum-dependent induction of cells into the S-phase). In the case of mock infections, cells were incubated with PBS and then washed following the same protocol as virus infections. Before the infection, one plate was kept in order to count the cells and calculated the virus volume in accordance to the following formula:

$$V_{virus} = MOI \times \text{total number of cells/virus titer}$$

40 MOI was used for *dl*1500 and *dl*312, 160 MOI for e1aRb$^-$, 60 MOI for e1ap300$^-$ and 6 MOI for e1ap400$^-$.

IMR90 cells were used at a passage lower than 10; HeLa cells were infected when reached a 60-70% confluency. After 24 hours the cells were detached from the plate by scraper for RNA extraction and western blot.

Construction of RNA-seq libraries and sequencing methodology

Cells were lysed with Trizol and total RNA extraction was performed with the Direct-zol RNA MiniPrep Plus (Zymo research). Total RNA-seq libraries of the two RNA-seq experiments (*dl*1500/*dl*312/mock infections or e1a binding mutant infections) were constructed following two slightly different methodologies. In the first case, an input of 1µg of total RNA was depleted of the ribosomal RNA using the Ribo-Zero rRNA Removal Kits (Epicentre). Total RNA libraries were afterwards constructed using the TruSeq stranded mRNA library Preparation kit (Illumina). A 100-base paired-end stranded sequencing was performed on a HiSeq4000 Illumina Sequencer using a sequencing depth of 60 million reads per sample.

For the construction of RNA-seq libraries of e1a binding mutants infected cells, 1µg of total RNA was treated with the RiboCop rRNA Depletion Kit (Lexogen). Total RNA libraries were constructed as previously described, whereas a 150 bp paired-end sequencing was performed on a NovaSeq sequencer (Illumina) using a sequencing depth of 100 million reads per sample.

RNA-seq data analysis and Alu profiling

Sequencing reads that passed the FastQC quality analysis were aligned to the GRCh38 human reference genome using STAR. Only uniquely mapped reads were considered in downstream analysis and subjected to counting with the HTSeq Python package. *Alu* pipeline was applied as in [147]. Read Count normalization and Differentially Expressed (DE) coding genes and *Alus* were performed using the DESeq2 package.

Only coding genes with a $|\log_2 FC| \geq 0.5$ and adjusted *p*-value ≤ 0.01 were considered as DE, whereas an adjusted *p*-value = 0.05 threshold was set to define significantly DE *Alu*s. Moreover, only *Alu* sequences detected in at least two replicates were considered for downstream analyses. Lowly expressed coding genes were filtered-out when the sum of DESeq2 normalized read counts among the two replicates was lower than 10.

GREAT analyses

A total of 1748 *Alu* sequences expressed in *dl*1500-infected cells and 656 *Alu* sequences expressed in *dl*312 infected cells were interrogated with the GREAT online tool (http://great.stanford.edu). The two nearest genes option was selected, imposing the limit to not exceed 100kb when defining a regulatory domain. Moreover, the curated regulatory domain option was also included. The output associated genes were afterwards intersected with expressed genes in our dataset.

Western blot

Protein extracts, separated by 9% SDS-polyacrylamide gel and transferred onto a PVDF membrane, were probed with antibody against e1a (anti-e1a MAb M73) and Ku86 H-300 (sc-9034; Santa Cruz), which was used as loading control. Anti-e1a MAb M73 was prepared as described previously [153]. e1a was detected with HRP-conjugated sheep anti-mouse IgG antibody (1:7500, BioRad), while Ku86 was detected with HRP-conjugated sheep anti-rabbit IgG antibody (1:7500, BioRad) and visualized with the Pierce ECL Western Blotting substrate (Thermo Fisher).

Real Time-PCR

cDNA was synthetized using the SuperScript III Reverse Transcriptase (Thermo Fisher) and Real Time reaction was assembled using the PowerUp SYBR Green Master Mix (Applied Biosystems). Real Time was performed using the Applied Biosystems 7500 Real-Time PCR System. The primers used for *Alu* and Pol III gene amplification

are listed in Table S2 of Supplementary Information I; the Ct values were normalized using the U1 snRNA gene as internal control.

ChIP-seq analyses

SigTags and poissP tags signals were retrieved from Ferrari et al. [86] and from Ferrari et al. [88]. Reads were aligned to expression-positive *Alu* loci after coordinate conversion (liftOver) from the hg19 to the hg38 human genome.

Wiggle files were created using the tool bamCoverage from the suite deepTools. Normalization was performed as counts per million (CPM) or as Reads Per Kilobase per Million mapped reads (RPKM). RPKM was only used when comparing H3K4me1 enrichment in *dl*1500- and mock-infected samples (Figure 16, bottom left panel).

The average ChIP signal was visualized using the tool Sitepro from the CEAS package (http://liulab.dfci.harvard.edu/CEAS/usermanual.html). SigTags were used for H3K27ac and p300, whereas poissP were used for H3K18ac and H4K4me1.

Chapter III

Gene Expression analysis on fibroblasts overexpressing *Alu* sequences

Results

As shown in Chapter II, e1a-induced *Alu* activation is a phenomenon limited to 0.1-0.2% of the intergenic/antisense *Alu* sequences, but the fold increase of specific *Alu* sequences is exceptionally elevated, with up to ≈73-fold increase of transcription compared to *dl*312. It is still not known if there are specific biological consequences for such activation, and if so, if *Alu* sequences are part of a cell stress response and/or are exploited by the virus to sustain its own replication. We asked therefore if the overexpression of *Alu* sequences in an unperturbed state of the cell has any effect on gene expression, and if *Alu* overexpression produces the same molecular phenotype as *dl*1500- infected cells.

Most of the results presented in this chapter is already published in Cantarella et al. [11].

Experimental strategy for Alu sequence overexpression

We overexpressed two *Alu* sequences in one cancer and one primary cell line (HeLa and IMR90 cells, respectively). The two *Alu* sequences *AluSq2* and *AluSx* were selected because already found reproducibly expressed in tumor and non-tumor cells. *AluSq2* is expressed in five ENCODE cell lines (H1-hESC, HeLa-S3, Hep G2, K562, NHEK) [31] and displayed a 32-fold induction in *dl*1500 infected fibroblasts (see results above). This *Alu* element is localized antisense to the first intron of the NFIA gene and lacks the canonical internal element A_5TACA_5, which is replaced by A_3G. *AluSx* is expressed in NHEK cells [31] and displayed a 2-fold increase compared to *dl*312 infected cells, as detected in our dataset (Chapter II). *AluSx* is antisense to the first intron of the AMFR gene and has a canonical A_5TACA_5 internal motif.

A portion of *Escherichia coli LacZ* gene antisense DNA sequence was used as a source of "random" DNA, which represented a negative control.

AluSq2 and *AluSx* were first amplified through PCR using human genomic DNA from saliva as a template. Since *Alu* elements are present in 1 million copy in the human genome and are highly repetitive, it is not possible to unambiguously amplify the *Alu* sequences of interest using primers that anneal at the 5′ and 3′ ends of the highly conserved sequences. Therefore, the genomic regions containing the target *Alu* elements were first amplified with primers that anneal to unique regions upstream and downstream the *Alu* sequences. The amplicon was afterwards used as a template for a second PCR reaction using primers that anneal at the 5′ and 3′ end of the *Alu* sequence, taking care to include a stretch of 4 Ts to ensure the presence of the Pol III terminator in the cloned sequence. The primers for the control sequence were designed in order to amplify a nucleotide sequence that have a similar length and GC distribution as the two *Alu* sequences (61% GC in *AluSq2*, 53% GC in *AluSx*, 61% GC in the Control sequence) (Figure 1).

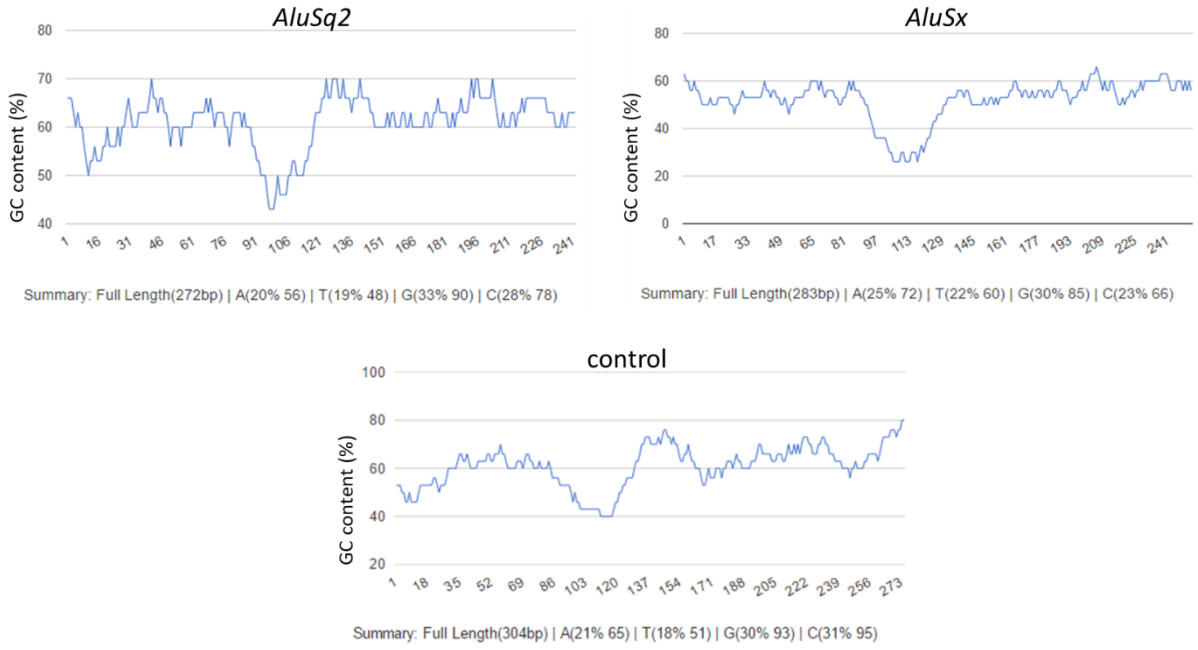


Figure 1. Evaluation of GC content in *AluSq2*, *AluSx* and control sequence. A decrease in the GC content in the centre of the nucleotide sequence is characteristic of each sequence analysed, resembling the nucleotide structure of a typical annotated *Alu* element.

AluSq2, *AluSx* and the control sequence were subsequently cloned into the pSUPER.GFP/neo vector (OligoEngine) under the control of the strong Pol III H1 promoter. Indeed, it is known that the internal *Alu* promoter does not drive the synthesis of high levels of *Alu* transcripts, and the presence of flanking sequences is necessary to obtain detectable RNA levels [36]. The RNase P RNA promoter (H1 promoter) is a type 3 Pol III promoter that was already shown to drive *Alu* transcription in transfected cells [154] and it is widely used for non-coding RNA overexpression studies [155]. Therefore, we chose this strategy to overexpress the two *Alu* elements and the control unrelated sequence from *E. coli*.

The nucleotide sequence and genomic coordinates of the overexpressed *Alu* elements and of the control sequence are available in Table S1 of Supplementary Information II, and the primers used in the cloning procedure are listed in Table S2 of Supplementary Information II.

In order to obtain the overexpression of the above described nucleotide sequences, the two *Alu* elements and the control sequence were subcloned from the pSUPER.GFP/neo vector into a 3rd generation lentiviral vector, as long as with an empty vector. The lentiviral vector also harbours the puromycin resistance gene and the eGFP gene, which allow the selection of stable transfectants and the monitoring of the integration through a fluorescent marker, respectively (Figure 2).

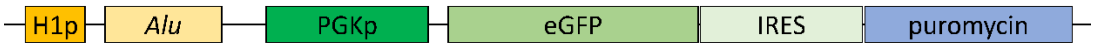


Figure 2. Schematic structure of the DNA inserted into the genome of IMR90 and HeLa cells to obtain *Alu* overexpression. Taken from [11].

IMR90 and HeLa cells were stably transformed with a lentivirus vector carrying *AluSq2*, *AluSx*, a control sequence or an empty vector. Stable integrants were selected for one week in the presence of puromycin and afterwards grown for other 7-10 days in the absence of the antibiotic. At the end of the selection it was possible to evaluate the presence of nearly 100% of GFP-positive cells, confirming the successful integration of the lentivirus construct (Figure 3).

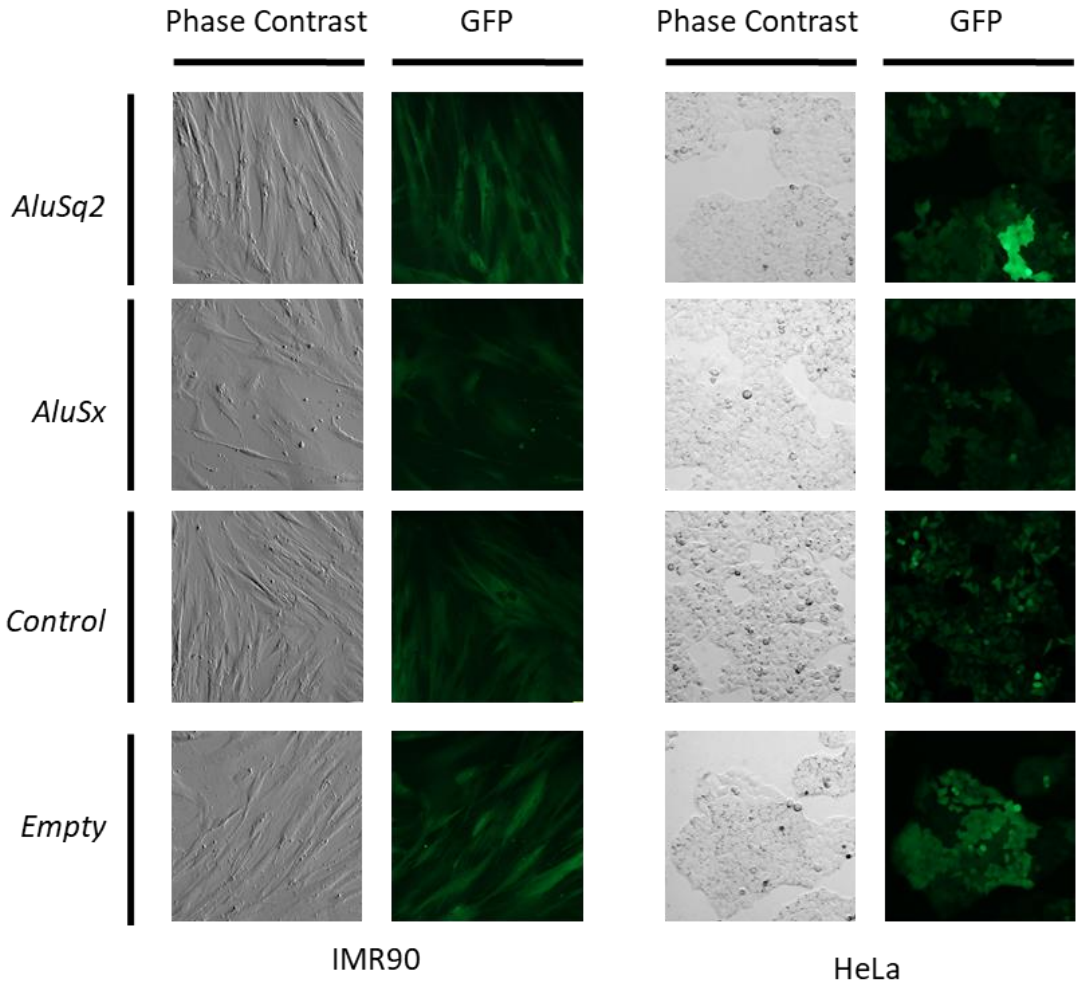


Figure 3. Integration of the lentivirus construct into the genome of IMR90 and HeLa cells. The majority of the cells were visualized as GFP-positive after 7 days of growth in the selective medium. The stable integration of *Alu* and control sequences, as well as the empty vector, could therefore be confirmed.

Validation of Alu and control sequences expression

Alus and control sequence overexpression was verified by RT-qPCR. In order to unambiguously detect *AluSq2* and *AluSx* sequences, the primers were designed on their unique 3' trailer region. A list of the primers used in this work is available in Table S2 of Supplementary Information II. RNA levels were normalized against the U1 snRNA gene, and the expression levels were calculated against cell lines transformed with an empty vector. As shown in Figure 4, *AluSq2* and *AluSx* have a 100-2000-fold increase in expression compared to cells transformed with an empty vector, which was used as a control. Interestingly, *Alu* overexpression in IMR90 cells was higher than in HeLa cells. Endogenous *Alu* RNAs are expressed at a background level in IMR90 and HeLa cells transformed with an empty vector. Since *Alu* sequences are known to be highly expressed in tumor cells compared to primary cell lines [141], the difference in *Alu* overexpression between the two cell lines could be explained with a higher background of *Alu* transcripts in HeLa cells transformed with an empty vector.

Since we were able to observe that the majority of cells were GFP positive and we were able to detect the over-expression of the H1p-controlled sequences, we could conclude that the lentivirus vector was efficiently integrated into the genome of IMR90 and HeLa cells.

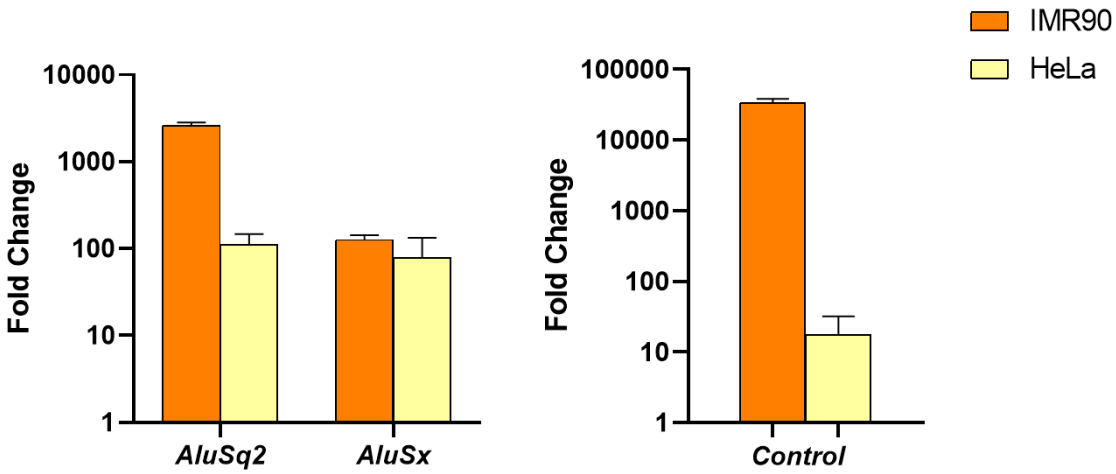


Figure 4. Validation of the overexpression of *AluSq2*, *AluSx* and the control sequence through RT-qPCR. *Alu* expression was evaluated using a set of primers that anneal to the unique region of the 3′ trailer, whereas the control sequence was detected with primers that anneal at the 5′ and 3′ end of the entire sequence. Gene expression was first normalized using the U1 snRNA gene as internal standard. Fold Change was calculated as a ratio between gene expression detected in *AluSq2*- *AluSx* or control-overexpressing cells and empty vector-transformed cells. Taken from [11].

Analysis of gene expression changes induced by Alu-overexpression

Changes in gene expression induced by *AluSq2*-, *AluSx*- and control-overexpressing cells were analysed through RNA-sequencing. Total RNA was extracted from IMR90 and HeLa cells and mRNA-sequencing was performed through poly(A) capture. Differentially Expressed (DE) genes were obtained comparing gene expression profiling of *AluSq2*/*AluSx*/control-overexpressing cells with cells transformed with an empty vector. Only genes with a $|\log2FC| \geq 0.5$ and an adjusted p-value < 0.001 were considered for downstream analyses. Interestingly, a total of 330 genes were found dysregulated in IMR90 cells, whereas only a few genes were dysregulated by *Alu* sequences in HeLa cells (Table 1). A list of the differentially regulated genes in IMR90 cells overexpressing *Alu* or control sequences is available in Table S3-S5 of Supplementary Information II.

Table 1. Number of up- and down- regulated genes in IMR90 and HeLa cells overexpressing *AluSq2*, *AluSx* or a control sequence.

	AluSq2		*AluSx*		Control	
	UP	DOWN	UP	DOWN	UP	DOWN
IMR90	87	101	147	105	86	55
HeLa	0	0	0	2	617	749

DE genes were detected with a highly significant adjusted *p*-value, and the most significant gene is represented by CCL2 in *AluSx*-overexpressing cells (adjusted *p*-value=2.50 x 10^{-97}) (Figure 4).

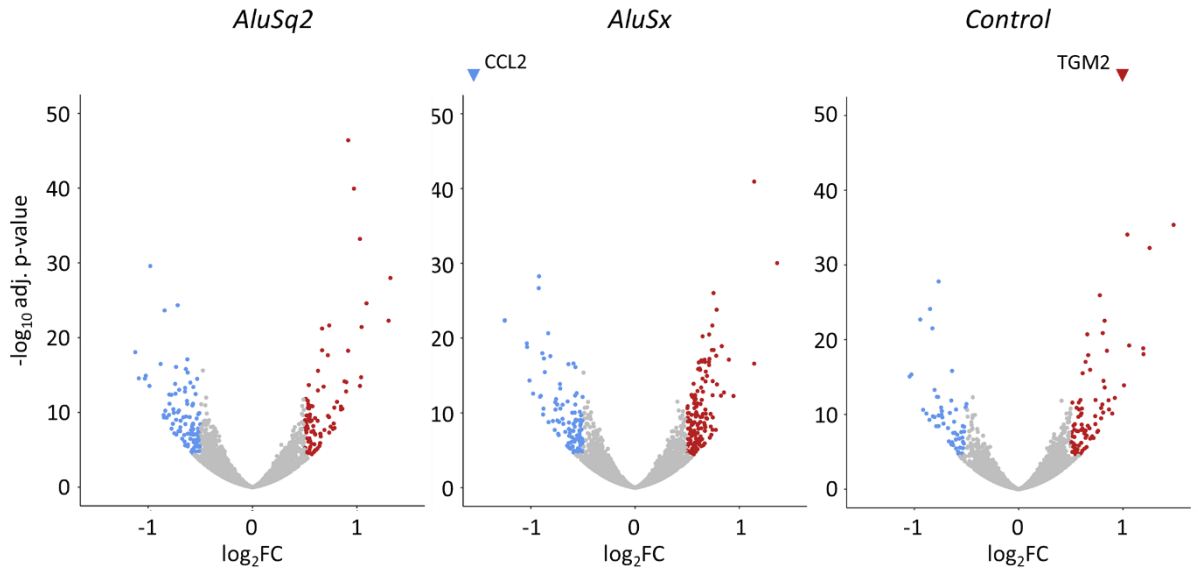


Figure 4. Volcano plot showing the DE genes in *AluSq2*-, *AluSx* and control-overexpressing IMR90 cells. DE genes show a |log2FC|≥0.5 and an adjusted *p*-value < 0.001. Upregulated genes are shown in red, downregulated genes are shown in blue. The CCL2 genes and TGM2 are out of scale, having an adj *p*-value of 2.50 x 10^{-97} (CCL2) and 3.41 x 10^{-53} (TGM2).

As shown in Figure 5, *AluSx* overexpression induced the dysregulation of the largest number of coding genes, whereas the control sequence modulated the expression of a lower number of genes compared to the two *Alu* sequences.

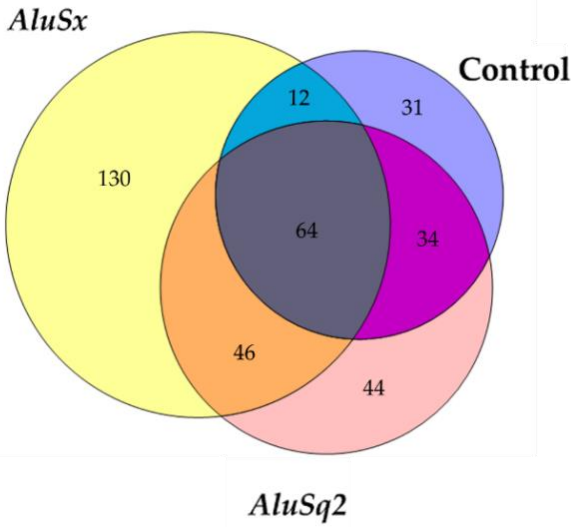


Figure 5. Venn diagram showing the number of genes that were detected as DE in IMR90 cells overexpressing *AluSq2*, *AluSx* or a control sequence.

We next proceeded in the examination of the dysregulated cell processes using the Ingenuity Pathway Analysis software. The most significant functions dysregulated by *Alu* overexpression are related to cell cycle, with *AluSx* inducing the most significant changes. Interestingly, pathways that promote cell cycle progression are predicted as activated (red bars), whereas pathways that are involved in an inhibition of cell cycle progression are predicted as inhibited (blue bars) (Figure 6). This suggests a role of *Alu* RNA as positive modulators of cell cycle progression. No pathways related to cell cycle were detected as significantly enriched in control-overexpressing fibroblasts, therefore we concluded that cell cycle progression is specifically activated by *Alu* sequences rather than abnormal levels of an exogenous RNA.

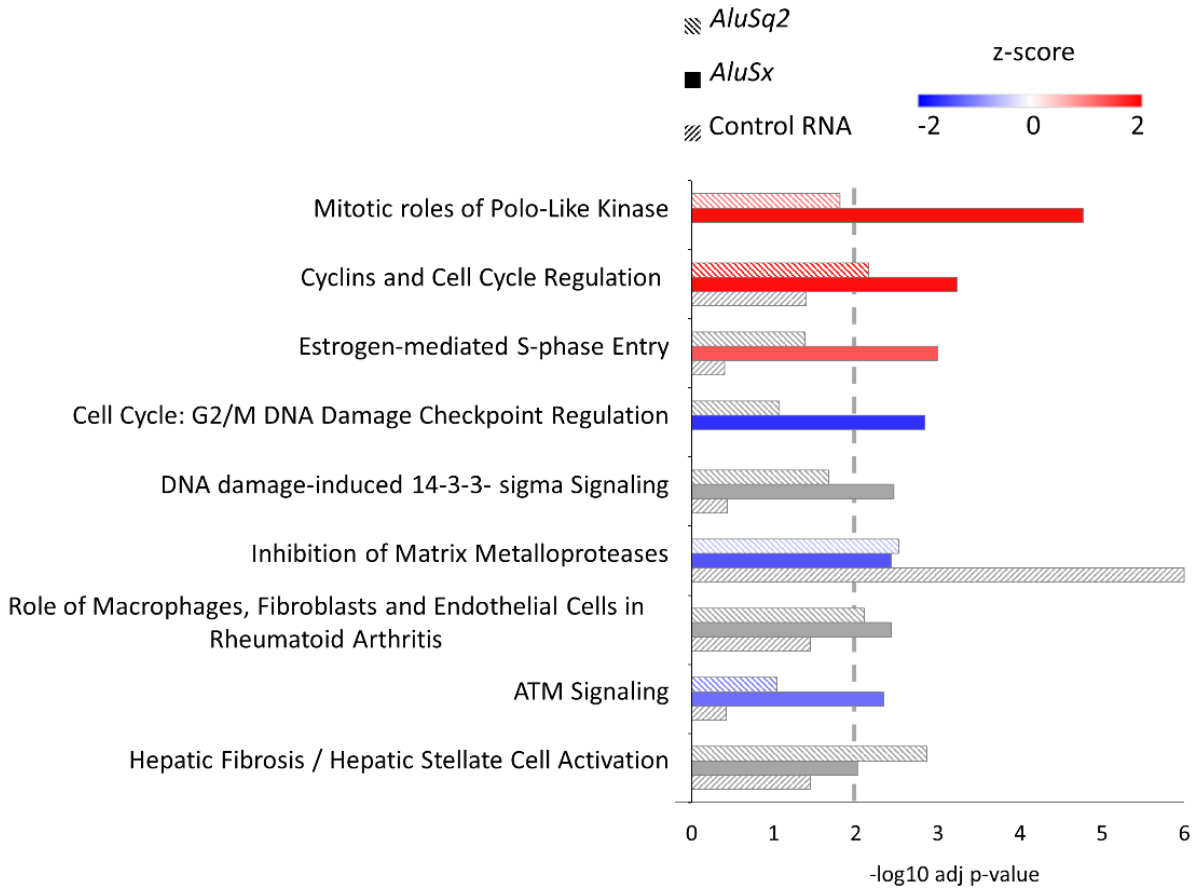


Figure 6. Cellular pathways dysregulated by the overexpression of *AluSq2*, *AluSx* and control sequences in IMR90 cells. The first four pathways are involved in the regulation of cell cycle. As indicated by the dashed grey line, these pathways were detected with an adjusted *p*-value (BH correction) < 0.01 in *AluSq2* or *AluSx* overexpressing cells, but no statistically significant enrichment was detected for control overexpressing cells. Enriched pathways were detected through analyses with the IPA software and are ordered by *AluSx* vs empty vector *p*-value. Activated or inhibited pathways are indicated by a positive (red) or negative (blue) z-score, respectively. Grey bars: no prediction can be made. Taken from [11].

Next, we performed prediction analyses to identify potential molecules (Transcription Factors or kinases) that could modulate DE genes that are related to mitotic pathways in *AluSx*-overexpressing fibroblasts. This prediction analysis is a feature of the IPA software, which allows to identify any molecular species (transcription factors, non-coding RNAs, kinases, drugs etc...) that can influence the transcription of a detected

DE gene in an RNA-seq input dataset. These so-called upstream regulators can be predicted to have a direct role in the modulation of DE genes or can be part of a cascade of other upstream regulators, leading ultimately to the modulation of the detected DE gene. Moreover, downstream phenotypes are also included in the analysis. All the regulatory relationships are based on experimentally validated molecular interactions reported in the literature.

In order to identify putative proteins that can mediate *AluSx*-modulation of cell cycle genes, we restricted the upstream regulator analysis on transcription factors and regulatory kinases that are known to control mitosis. We were able to identify the transcriptional co-activator YAP-1, the transcription factor FOXM1 and the cyclin-dependent kinase inhibitor 1 (CDKN1A) as direct DE genes regulators (Figure 7). It is not clear how *AluSx* can modulate YAP1 and CDKN1A action, since these genes were not detected differentially expressed in our dataset. However, the case of FOXM1 is particularly interesting, since it is also upregulated by *AluSx* (Figure 8). The genes modulated by the upstream regulators predicted by IPA analyses belong to CDC proteins, cyclins and cyclin-dependent kinases (CCN and CDK proteins), proteins involved in mitotic spindle assembly and chromosome segregation (CENPA, KIF20A, PLK1, BUB1B), and signalling molecules that are involved in tumorigenesis when their expression is dysregulated (CYR61, EDN1, SFRP1, PDGFB).

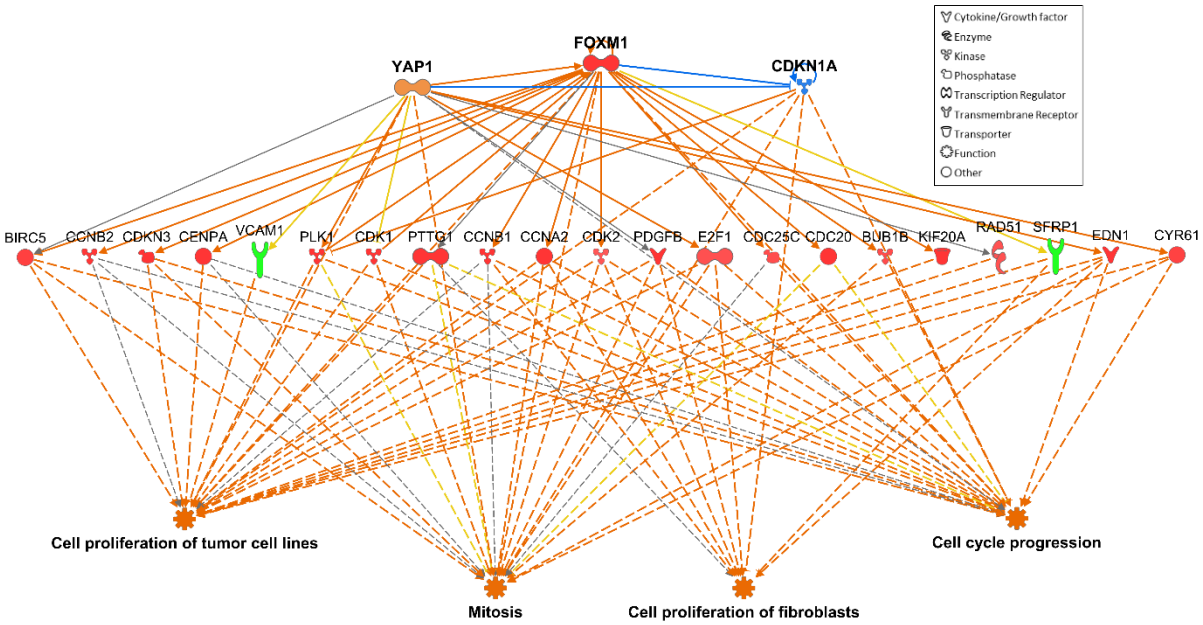


Figure 7. Possible upstream regulators of DE genes that belong to cellular pathways involved in mitosis. Upstream regulators are represented in the upper part of the picture, DE genes detected in *AluSx* and cell phenotypes are shown in the middle and bottom part of the figure, respectively. Regulatory relationships between upstream molecules, the DE genes and cell functions are predicted as inhibitory (blue lines) or activating (orange lines), whereas yellow lines indicate an inconsistent prediction; grey lines are inserted where no predictions could be made. Continuous and dashed lines show direct and indirect interactions (less than three passages), respectively.
Red symbols=upregulated genes. Green symbols=downregulated genes. From [11].

Interestingly, eleven out of the 21 DE genes belonging to the network in Figure 7 are also dysregulated in *AluSq2*-overexpressing cells, whereas only four genes are differentially expressed in control-overexpressing cells (Figure 8).

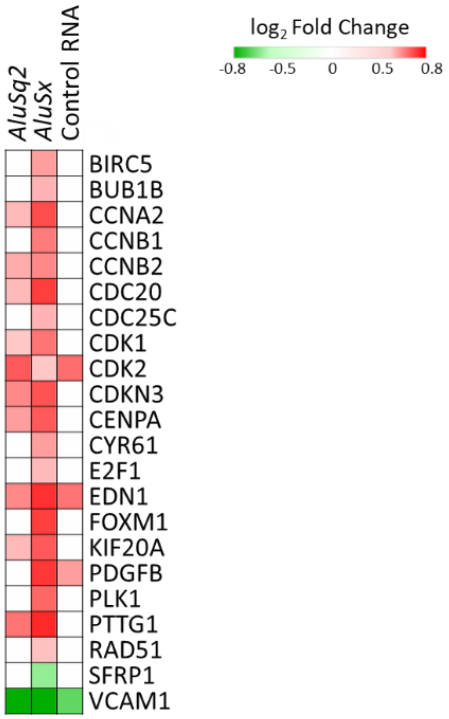
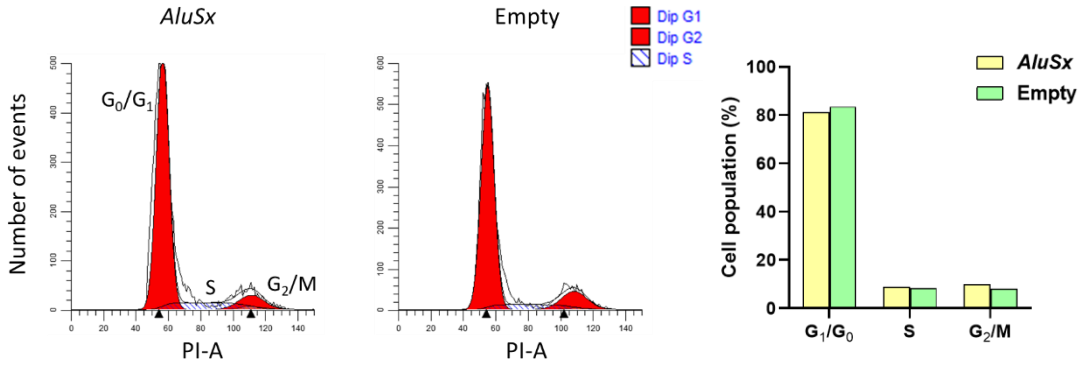


Figure 8. Heat map showing DE genes belonging to cell pathways related to mitosis. The log_2 Fold Change of DE genes shown in the middle part of Figure 7 are shown for *AluSq2*, *AluSx* and control overexpressing fibroblasts. From [11].

Alu RNA induces cell cycle progression into the S-phase

In order to experimentally verify the *Alu*-stimulation of cell cycle that was predicted from Differential Expression analyses, we assessed the progression into cell cycle through flow cytometry. Cell division was first inhibited with incubation in serum free medium for 24h and cell cycle arrest was detected through FACS analyses. As shown in Figure 9A and 9B, serum withdrawal was effective in inducing a cell cycle arrest in G_1/G_0 of 83 ± 2% of cells, and around 8% of cells were detected in S-phase and G_2/M phase.



(A) (B)

Figure 9. Cell cycle arrest of *AluSx* and empty vector-transformed fibroblasts. Serum starvation for 24 hours resulted in the accumulation of transformed fibroblasts in the G_0/G_1 phase, as revealed by flow cytometry analyses. DNA content was evaluated through PI staining and cell cycle distributions were visualized through the ModFit software (A)**.** The percentage of cells in G_0/G_1, S or G_2/M phase is shown by the bar plot in B.

After serum starvation for 24 hours, cell cycle was induced by adding fresh medium with 10% serum and cell cycle was allowed for other 24 hours of cell culture. Afterwards, cell cycle distribution was analysed by flow cytometry. We took care of including two independently transformed cell lines for each sample, and each transformation was analysed for one-to-three different cell plates, in order to obtain a sample size between four and six replicates for each *Alus*/control/empty-overexpressing cell lines. As shown in Figure 10, *AluSq2*- and *AluSx*- overexpressing cells have a significant higher percentage of cells in S-phase and a lower percentage of cells in the G_2/M phase compared to control and empty-transformed cells.

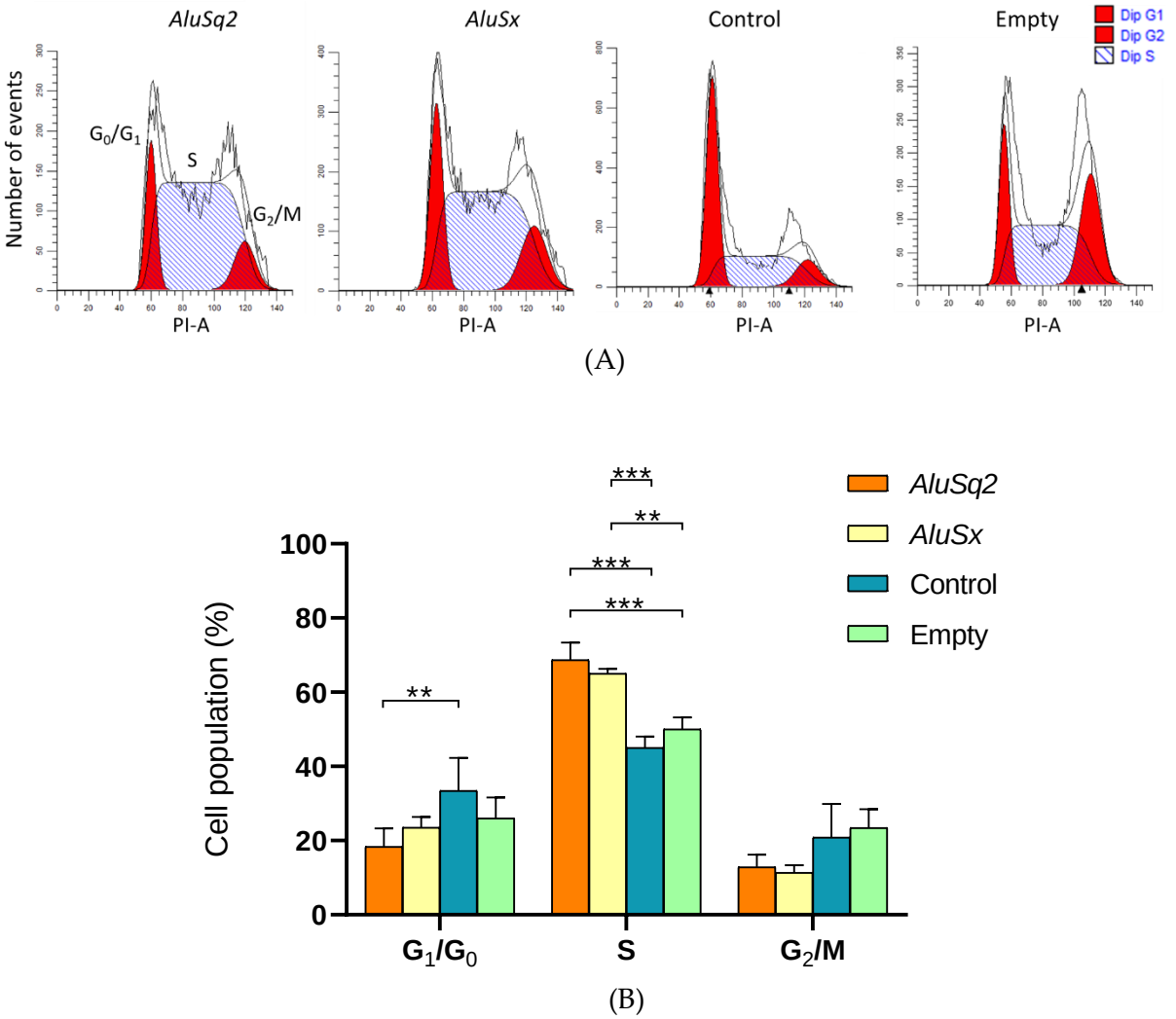


Figure 9. *Alu* sequences stimulate cell cycle arrested fibroblasts to enter the S-phase. (A) Cell cycle distributions of *AluSq2*, *AluSx*, control and empty vector-transformed fibroblasts were visualized with the ModFit Software. Cell cycle arrest was performed by serum withdrawal for 24 hours, and cell cycle was released by adding 10% serum fresh medium for other 24 hours. DNA content was detected through PI staining. (B) Bar plot showing cell populations that belong to the G_0/G_1, S and G_2/M phase. *p*-values were calculated as a 2-tailed *Student's t-test* (equal variance) or a *Welch's t-test* (unequal variance). Cell numbers from each sample were obtained from at least four replicates, which were obtained from two independently transformed IMR90 cells. *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001. From [11].

Predicted molecular mechanisms of Alu-mediated gene regulation

As described in Chapter I, retrotransposons can regulate gene expression through a variety of mechanisms, among which Staufen1-mediated mRNA decay and miRNA-like sponge mechanisms have been hypothesized.

Staufen1 can be recruited by dsRNA formed by intermolecular base pairing between two RNA molecules, triggering Staufen1 (STAU1)-mediated mRNA decay (SMD). *Alu* sequences embedded in Pol II transcripts are known to form intermolecular base pairing, recruiting STAU1 and therefore mediating mRNA degradation [125,126]. We reasoned that the overexpressed free *Alu* RNA (*AluSq2* and *AluSx*) could base-pair with mRNAs that have *Alu* sequences embedded in their 3′ UTR, with a similar mechanism as proposed by Gong et al. [126], and Gong et al. [125]. Therefore, we analysed the presence of *Alu* sequences in the 3′ UTR of DE genes detected in *AluSq2* and *AluSx*-overexpressing fibroblasts.

As shown in Table 2, the percentage of genes in *AluSq2*-overexpressing cells harbouring an *Alu* element in its 3′ UTR significantly increased in the case of down-regulated genes (36% vs 24%), whereas it decreased when considering upregulated genes (10% vs 24%). Similarly, the presence of 3′ UTR *Alu* in up-regulated genes of *AluSx*-overexpressing cells significantly decreased to 17%, whereas the difference in percentage was not statistically significant in the case of down-regulated DE genes. Based on these data, a mechanism of base pairing between an *Alu* embedded in the 3′UTR and the free overexpressed *Alu* could be hypothesized. In support of this speculation, we verified an enrichment of antisense *Alus* in the 3′ UTR of downregulated genes (55%) compared to the totality of the genes in the genome (49%) in *AluSq2*-overexpressing cells (Table 2). A list of all the *Alu* sequences detected in the 3′ UTR of DE genes is available in Table S6 of Supplementary information II.

Table 2. Number of DE genes that have an *Alu* sequence embedded in their 3′ UTR region. The minimum length of *Alu* spanning the 3′ UTR region was set as 100-nt: *Alu* sequences that are present for less than 100-nt in the 3′ UTR were not considered in the analysis.

Sample	Genes with *Alu* in 3′ UTR	Total Number of Analyzed Genes	*p*-Value (Fisher exact test)	% of Genes with an *Alu* Element in Their 3′ UTR	% of Anti-sense *Alu* in Each Gene (Average)
Genome	4838	19,836	-	24.39	49.22
*AluSq2*_Up	9	87	0.0005	10.34	66.67
*AluSq2*_Down	37	101	0.0019	36.63	55.63
*AluSx*_Up	25	147	0.0082	17.01	30.00
*AluSx*_Down	28	105	0.0761	26.67	68.45

From bioinformatic analyses, it is known that *Alu* elements can be targeted by a vast multitude of miRNA sequences [156], increasing the possibility of *Alu* involvement in the miRNA-based gene regulation network. Moreover, it has been proposed that miRNAs can be influenced in their regulatory properties by abundant long non-coding RNAs (lncRNA) carrying a miRNA Responsive Elements (MRE). miRNA binding to MRE within a lncRNA would preclude miRNA-targeting of a similar MRE within an mRNA molecule, hindering miRNA-directed degradation of the coding transcript [157,158].

Therefore, we searched putative miRNA response elements embedded in *AluSq2* and *AluSx* sequences interrogating the databases miRbase and miRDB. We considered as potentially targeting miRNAs only those sequences that have an Expected value < 0.05 in miRbase or a Target score > 80 in miRDB. No significant miRNA response elements were found on *AluSq2*, whereas the sequences *hsa-miR-619-5p* and *hsa-miR-7151-3p* were detected as likely targeting *AluSx*. We next analysed the DE genes that could be targeted by these two miRNAs. Interrogating the database miRDB, we found that each miRNA could target two DE genes in *AluSx*-overexpressing cells (Table 3).

Interestingly, both miRNAs were found to target the gene PRR11, which is slightly up-regulated in our gene dataset and is reported to be implicated in cell cycle progression and lung cancer [159].

Table 3. miRNAs targeting the *AluSx* sequence and DE genes detected in *AluSx*-overexpressing fibroblasts. *hsa-miR-619-5p* and *hsa-miR-7151-3p* were detected from miRbase and miRDB as potentially targeting *AluSx*, respectively. Potential DE gene targets of these two miRNAs were retrieved from miRDB.

miRNA	Gene target	log2 Fold Change
hsa-miR-619-5p	PRR11	0.554
	MASP1	-1.036
hsa-miR-7151-3p	PRR11	0.554
	SVEP1	-0.834

Overall, the data support a mechanism of gene regulation based on intermolecular base pairing between *AluSq2* and an *Alu* embedded in the 3' UTR of DE genes, whereas *AluSx*-mediated upregulation could be dependent on a miRNA-like sponge mechanism (Figure 10).

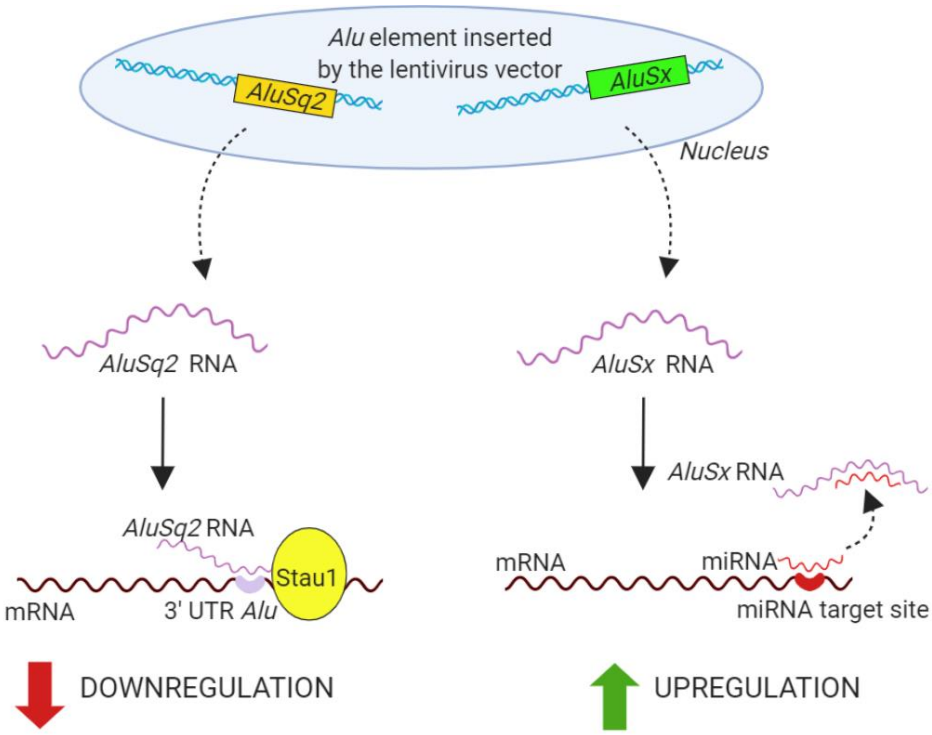


Figure 10. Putative mechanism of post-transcriptional gene regulation by *Alu* RNAs. *AluSq2* free RNA could lead to downregulation of genes with an *Alu* element embedded in their 3′ UTR: double-stranded RNA formed by base pairing between the two *Alu* species could recruit the STAU1 protein, triggering mRNA degradation. On the other side, gene silencing could be mediated by *AluSx* through miRNA sponging. *AluSx* RNA and mRNAs could bear similar miRNA response elements (MRE). Therefore, miRNA-mRNA targeting could be hindered by the recruitment of the same miRNA species on the *AluSx* transcript, inhibiting miRNA-mediated mRNA decay resulting in gene upregulation.

Materials and Methods

Amplification of Alu and control sequences from genomic DNA

The *AluSq2* sequence was obtained through PCR amplification using human genomic DNA from saliva as a template. The primers *AluSq2*_chr1_fw and *AluSq2*_chr1_rev were designed to anneal 200 nucleotide upstream the *Alu* TSS and 100 nucleotide downstream the Polymerase III terminator, respectively. The amplicon was first cloned into the pGEM-T easy vector (Promega) and subsequently amplified by PCR reaction using the primers *AluSq2*_pSUPER_Bgl_fw and *AluSq2*_pSUPER_Xho_rev. This strategy ensures to unambiguously amplify the *AluSq2* sequence without contamination from any of the 1 million *Alu* repetitive sequences in the human genome. Moreover, the introduction of the restriction site *BglII* at the 5′ end and *XhoI* at the 3′ end allows the subcloning of the *AluSq2* sequence into the pSUPER.basic vector (OligoEngine).

The *AluSx* sequence was isolated from human genomic DNA using a more direct strategy, involving two PCR reactions and only one cloning passage. A first PCR was performed on saliva genomic DNA using the primers *AluSx*_Fw and *AluSx*_Re, which unambiguously anneal 183 nt upstream the *Alu* TSS and 75 nt downstream the Pol III terminator, respectively. The resulting amplicon was used as a template for a second PCR reaction, obtaining the *AluSx* sequence carrying the restriction start sites *BglII* and *XhoI* at the 5′ and 3′ ends, respectively (primers *AluSq2*_pSUPER_Bgl_fw and *AluSq2*_pSUPER_Xho_rev).

Lastly, the control sequence was isolated from *Escherichia coli* LacZ gene (*Escherichia coli* DH10 genomic DNA) through a two-step PCR strategy. A first PCR was performed using the primers LacZ_fw, which introduces the restriction site BglII at the 5′ end, and LacZ_re2. A second PCR was then performed using the same primer forward LacZ_fw and the reverse primer LacZ_re3, which introduces the typical poly(A) tail

of *Alu* sequences and the restriction site *HindIII* that is exploited for subcloning into the pSUPER.GFP/neo vector.

Construction of lentiviral vector

The *Alu* sequences were first cloned into the pSUPER.GFP/neo (OligoEngine) vector using the restriction enzymes *BglII* and *XhoI*, in order to insert the sequences downstream the H1 promoter (H1p). During the cloning procedure, we noticed that the *AluSx* sequence harbours a *HindIII* restriction site 13 nt upstream the Pol III terminator. Therefore, we decided to exploit *HindIII* site to clone the control sequence upstream the Polymerase III terminator.

The pSUPER.GFP/neo vectors carrying *AluSq2*, *AluSx* and the control sequence were used as a template to isolate H1p-*Alu*/control sequences that were cloned afterwards into the lentiviral vector. Four different PCR reactions were performed using a common forward primer 5'_XbaI_H1_prom and the specific reverse primers 3'_BamHI_*AluSq2*, 3'_BamHI_*AluSx*, 3'_BamHI_control and 3'_BamHI_no_insert, which was used for the construction of an empty vector that was used as negative control. H1p-*AluSq2*, H1p-*AluSx*, H1p-control and H1p-empty sequences were cloned into a 3rd generation lentiviral vector pRRL-MCS-PGK-GFP-IRES-Puro that was pre-digested with *XbaI* and *BamHI* restriction enzymes. The clones carrying the correct inserts, as verified by DNA sequencing, were amplified and purified using an Invitrogen PureLink HiPure Maxiprep Kit.

Lentivirus particles encoding *AluSq2*, *AluSx*, a control sequence and an empty vector were generated by transient co-transfection of 293T cells with a four-plasmid combination system, as described in Naldini et al. [160], with slight modification. The VSV-G viral envelope was produced from the pMD.G construct, the pMDLg/pRRE and pRSV-REV plasmids were used as packaging constructs, while the pRRL plasmid correspond to the transfer vectors. The co-transfection of the four plasmids was performed in 100 mm dishes of non-confluent 293 cells by the CaPi-DNA

coprecipitation method [161,162]. After 48h the conditioned medium was harvested and passed through 0.45 mm filters. p24 antigen concentration was assessed by ELISA (the Alliance HIV-I p24 ELISA Kit) to determine viral titer, which is expressed as μg of p24 equivalent units per milliliter.

Generation of stable cell lines

IMR90 and HeLa cells (ATCC) were grown in Dulbecco's Eagle Modified Medium (Thermo Fisher) supplemented with 10% FBS, 100 U/mL penicillin/streptomycin (Thermo Fisher) and maintained at 37 °C in a 5% CO_2 atmosphere. Exponentially growing IMR90 and HeLa cells were plated in 6-well plates at a density of 2×10^5 cells per well. The next day, the medium was aspirated and cells were transduced with 1ml of lentivirus (0.86 μg/mL *AluSq2*, 0.75 μg/mL *AluSx*, 1.2 μg/mL control, 0.89 μg/mL empty vector) supplemented with 4 μg/mL protamine sulfate. After overnight incubation the cells were washed with fresh medium and grown for eight days before starting the antibiotic treatment to select stable integrants. Cells were grown for one week in puromycin selective medium (2 mg/ml for IMR90 cells and 0.5 μg/mL for HeLa cells), and afterwards for other 7-10 days to allow cell expansion before RNA extraction. The experiment was performed in duplicate.

Validation of AluSq2 and AluSx overexpression by Real-Time PCR

Total-RNA was extracted from exponentially growing HeLa cells and passage 9 IMR90 cells (Direct-zol RNA MiniPrep Plus kit, Zymo Research), and the cDNA was synthesized using the SuperScript III Reverse Transcriptase (Thermo Fisher). Real Time reactions were assembled using the PowerUp SYBR Green Master Mix (Applied Biosystems) and gene expression was assessed using the Applied Biosystems 7500 Real-Time PCR System. A pair of previously optimized primers that anneal to the 3' trailer of *Alu* sequences was used to detected *AluSq2* (*AluSq2*_qRT_fw and *AluSq2*_qRT_rev) and *AluSx* (*AluSx*_qRT_fw and *AluSx*_qRT_rev) overexpression.

The overexpression of the control sequence was evaluated using the same primers as used in the cloning strategy (Primers LacZ_fw and LacZ_re2). The Ct values were normalized against the Ct values obtained from U1 snRNA, which was used as internal normalization control. All the primers used in Real Time PCR reactions are listed in Table S3 of Supplementary Information II.

Construction of RNA-seq libraries and Differential Gene Expression analyses

mRNA libraries were prepared from 1µg of total RNA using the TruSeq stranded mRNA library Preparation kit (Illumina) and a 50 base-pair stranded single end sequencing was performed on a HiSeq4000 Sequencer (Illumina). Reads were aligned to the GRCh38 human reference genome using STAR [163], gene counts were obtained by HTSeq [164] and the differential gene expression analyses were performed through DESeq2 [165]. Genes with a $|\log2FC| \geq 0.5$ and a adjusted *p*-value ≤ 0.001 were considered as Differentially Expressed and visualized using the Volcano Plot workflow on the Galaxy web platform at the public server at https://usegalaxy.eu. Venn diagram was visualized with Venn Diagram Plotter software (https://omics.pnl.gov/software/venn-diagram-plotter), and pathway enrichment analyses and upstream regulators analyses were performed through the Ingenuity Pathway Analysis software (IPA, QIAGEN). Only pathways with overlap p-values (BH-adjusted) ≤ 0.01 were considered significantly enriched. HeatMaps were visualized with Morpheus software (https://software.broadinstitute.org/morpheus).

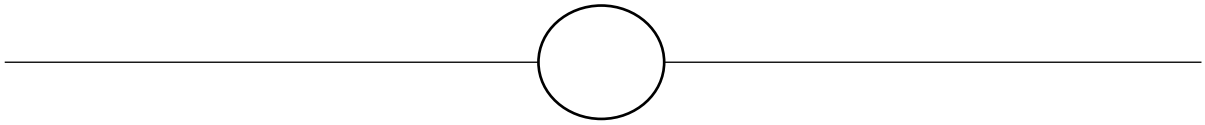
Cell cycle analyses by flow cytometry

In order to detect a possible *Alu*-induction into the S-phase, cells were first synchronized with serum starvation for 24 h and then cycling was allowed for other 24 h before flow cytometry. Briefly, 5×10^5 cells were seeded in 100 mm dishes in growth medium overnight. The next day, cells were washed with PBS and incubate with serum free medium for 24 h, at the end of which the medium was switched with

10% FBS medium to release cell cycling. After 24 h, cells were detached by trypsinization and fixed with 70% ethanol in PBS for at least 24 h at -20 °C. On the day of flow cytometry analyses, cells were washed with PBS and resuspended in propidium iodide (PI) staining solution (100 μg/mL PI, 20 μg/mL RNase). After an incubation time of 15 minutes at room temperature, cells were analysed using a BD FACS Celesta flow cytometer (BD Biosciences). 20000 cells were analysed from at least four replicates, which were obtained from two independent transformations of IMR90 cell lines transformed with the lentiviral vectors. The ModFit LT software (version 5.0, Verity Software House) was used to analyse cell cycle distributions. An F-test was performed to assess equal variances for *Alu*Sq2/*Alu*Sx vs control/empty comparisons for each cell cycle phase. Unequal variances were detected in *Alu*Sx vs control in G_0/G_1-phase, *Alu*Sx vs empty in the S-phase, *Alu*Sq2 vs control in G_2/M-phase. In these cases, the *p*-value was calculated according to a *Welch's* t-test. A *Student's* t-test was used to perform statistical analyses for all the other comparisons.

Evaluation of genomic Alu location and miRNA enrichment

The presence of *Alu* sequences in the 3′ UTR of DE genes was analysed using GENCODE v27 as a reference database, taking into account only *Alu* sequences that had a minimum of 100 nt of overlap with the 3′ UTR of a Differential Expressed gene. miRNA target elements on *AluSx* were detected using the databases miRbase and miRDB, considering as significantly enriched only those miRNAs with an E-value < 0.05 (in miRBase) and a Target Score > 80 (in miRDB). *hsa-miR-619-5p* and *hsa-miR-7151-3p* target genes were retrieved using miRDB (target score > 80) and matched with DE genes of *AluSx*-overexpressing fibroblasts.



Chapter IV

Discussion

The molecular mechanisms of *Alu* silencing in physiological conditions started to be elucidated in the last decades. However, potential factors that induce a derepression in response to virus infection and the functional implications of *Alu* overexpression are still unknown. This study analyses the expression of *Alu* elements under a mechanistic and functional approach.

First, *Alu* expression profiling was performed in cell cycle arrested fibroblasts infected with Adenovirus 5 only expressing the oncogenic small E1A protein (*dl*1500). e1a induces G_0/G_1 fibroblasts to enter the S-phase [86-88], therefore the analysed *Alu* response was in the context of an enhanced cell cycling. *Alu* elements are very lowly expressed in unstimulated cells. Indeed, we could detect only 648 expression-positive *Alu*s in response to *dl*312 infection and 669 *Alu*s in unperturbed cells (mock-infections), originating from roughly the 0.1% of annotated intergenic/antisense *Alu* loci.

*dl*312 mutant virions are unable to produce any viral protein. Therefore, the comparison between *dl*1500 and the additional negative control *dl*312 allows to distinguish between *Alu*s potentially activated by non-specific processes, such as virion internalization, from processes uniquely related to e1a expression. Remarkably, the number of expression-positive *Alu*s increases almost three times in *dl*1500-infected cells compared to control infections.

Moreover, we could detect a 4-fold increase in *Alu* average expression in e1a infections compared to *dl*312/mock controls (Figure 4A of Chapter II), with over-expression up to 73 folds in single *Alu*s. This fold change discrepancy suggests that *Alu* transcription is a variable phenomenon, which could possibly depend on a multitude of factors that are characteristic of each *Alu* locus (i. e. nucleotide sequence, chromatin environment and/or the presence of epigenetic marks). Importantly, e1a could enhance the transcription of *Alu*s already expressed in unstimulated cells (called common *Alu*s), in addition to the activation of newly expressed *Alu*s in *dl*1500 that were absent in *dl*312/mock infections (activated *Alu*s) (Figure 3, Figure 4B and Figure 5 of Chapter II). Interestingly, common *Alu*s showed higher read counts compared to newly activated

*Alu*s in *dl*1500-infected cells (Figure 5), suggesting an efficient transcriptional enhancement of those sequences that are already expressed at a basal level before e1a stimulation. This could be explained with the existence of a pre-marked state that allows a basal transcription in unperturbed cells. e1a could exploit this feature, consequently inducing a higher transcription of pre-marked *Alu*s. However, it is likely that other mechanisms might influence *Alu* overexpression, since the induction of not pre-marked *Alu*s is also observed.

It is possible that *Alu* expression could be influenced by their genomic insertion in regions of actively transcribed genes. Indeed, more than 80% of expression-positive *Alu*s in *dl*1500- and *dl*312-infected cells are located within 100-kb upstream or downstream the Transcriptional Start Site (TSS) of expressed genes in our dataset. Even though we could verify the genomic association of expressed *Alu*s with expressed genes, the functional significance of this association is not clear. One possibility might be the induction of chromatin changes at target gene regions that also host *Alu* elements, leading to the indirect activation of *Alu* expression. A more fascinating hypothesis is based on recent evidence of *Alu* function as enhancer elements [67]. Enhancer-promoter interactions through chromatin looping can take place in chromatin sub-domains containing approximately 100-200 kb of DNA [166]. Moreover, recent evidence show that enhancer SINEs can also be transcribed in enhancer RNA (eRNA) and promote the association of Pol III with the Pol II machinery to induce gene expression [122]. Based on this evidence, it is tempting to speculate that *Alu* elements function as enhancer of gene expression through chromatin looping, which would be facilitated by *Alu* transcripts arising from e1a stimulation.

Our studies on e1a binding mutants identifies p400 as an e1a interactor required to enhance *Alu* transcription, since in the absence of e1a binding to p400 *Alu* expression did not increase to similar levels as in wt e1a. Strikingly, we observed an enrichment of the histone mark H3K4me1 at expression-positive *Alu*s in both *dl*1500- and mock-infected cells. It is well established that active enhancers are marked by H3K4me1 [66],

which can be recognized by the chromodomain of the TIP60 transcriptional co-activator [110]. Therefore, we hypothesize that the interaction of e1a with p400 might stabilize the p400-TIP60 complex, with the consequent recruitment to *Alu* loci marked by H3K4me1 and ultimately induction of *Alu* transcription. This mechanism is in agreement with a selective enhancement of *Alu* transcription, as the expression of other Pol III genes (7SL and U6) did not change in response to wt e1a and e1ap400$^-$ expression.

The transcription factor c-Myc might also be involved in *Alu* overexpression mediated by e1a-p400 binding. Indeed, e1a-p400 interaction is known to stabilize c-Myc [117], which in turn can enhance Pol III transcription through association with TFIIIB, as reported by Gomez-Roman et al. [167].

Contrarily to the evident requirement of e1a-p400 interaction to enhance *Alu* expression, the contribution of Rb and p300 is not equally clear. The slightly lower expression of *Alu* sequences in e1aRb$^-$ compared to wt e1a (Infection1 of Figure 13, Chapter II), could be explained by the incapacity of e1a to interact with Rb, maintaining Rb repression of TFIIIB [75].

Lastly, ChIP-seq data showed no enrichment of p300 and its histone acetylations at expression-positive *Alu*s, therefore a different mechanism than chromatin remodeling should be taken into consideration to explain reduced *Alu* activation in e1ap300$^-$. The tumor protein p53 is a known target of p300 acetylation [168] and a Pol III suppressor through an inhibitory interaction with TFIIIB [169]. We speculate that e1a binding to p300 could favour p53 acetylation, resulting in an inhibition of p53-TFIIIB association. Therefore, the incapacity of e1ap300$^-$ to overexpress *Alu* sequences might be explained by failing of p53 acetylation, without removing p53 inhibitory effect on TFIIIB resulting therefore in a less efficient *Alu* overexpression.

Overall, the lack of binding of e1a with the host proteins does not lead to a complete abrogation of *Alu* expression. Therefore, it is possible that a simultaneous interaction of e1a with Rb, p300 and p400 is required to stimulate *Alu* expression. In addition, it

can not be excluded that other e1a interactors could be involved in the modulation of *Alu* expression.

The functional significance of *Alu* overexpression after e1a infection is an unsolved biological issue. *Alu* RNA could take part in the cellular immune response to counteract virus infection; on the opposite, *Alu* overexpression might be one among many mechanisms used by e1a to drive the cell to enter the S-phase. Interestingly, our data show that growth-arrested fibroblasts are induced to cell cycling by virus infection or by the overexpression of exogenous *Alu* sequences.

Alu overexpression was obtained through the establishment of stable cell lines. This ensures an experimental set up that is more similar to a physiological cell environment, compared to transient transfection with abnormal levels of an expression plasmid or RNA. However, this approach ensured an efficient overexpression of the *Alu* sequences *AluSq2* and *AluSx* in primary (fibroblasts) and tumor (HeLa) cell lines. Indeed, we could observe approximately 2500-fold and 110-fold increase of *AluSq2* in fibroblasts and HeLa cells, respectively. *AluSx* was overexpressed 126-fold and 78-fold in fibroblasts and HeLa, respectively. The lower fold induction observed in HeLa cells might be due to a higher expression of endogenous *Alu* sequences compared to fibroblasts, consistent with the notion that *Alu* RNAs are highly expressed in tumor cells [141].

Alu overexpression did not induce any gene expression changes in HeLa cells, whereas a modulation of cell cycle genes was observed in fibroblasts. In general, *Alu* RNA induced a slight modulation of gene expression, with only a few genes being at least 2-fold up- or down- regulated (Figure 4 and Figure 8 of Chapter III). This could explain the failure in detecting a differential expression of cell cycle genes in HeLa cells, which are known to be already proliferatively dysregulated [170].

The inclusion of an *Alu*-unrelated sequence in our experimental set up allowed to discriminate between effects that are *Alu*-specific or merely due to the overexpression of an exogenous sequence. Indeed, we found a subset of genes that is modulated by

an unrelated control sequence. *AluSx* seemed to have the strongest effect on gene regulation compared to *AluSq2*, given the higher number of differentially expressed genes (Figure 5 of Chapter III) and of significantly dysregulated pathways (Figure 6 of Chapter III). This dissimilarity could be explained by the different nucleotide sequences of the two overexpressed *Alu* elements, since *AluSq2* lacks the internal A-rich region that separates the left and the right arm of the *Alu* dimer. The two *Alu* RNA arms fold independently while being connected by an internal flexible linker [171]. The lack of this region in *AluSq2* might result in a different RNA folding, therefore possibly leading to different effects on gene regulation.

Strikingly, cell cycle of human fibroblasts was especially induced by *AluSx* overexpression, as revealed by RNA-seq and flow cytometry analyses. The induction of cell cycle in primary fibroblasts was relatively unexpected. Indeed, *Alu* overexpression through transient transfection is known to inhibit the proliferation of HeLa cells [172], stem cells [173] and prostate cancer cells [143]. Moreover, a decrease in viability and induction of pro-apoptotic changes was previously observed in breast adenocarcinoma cells [174], whereas *Alu* overexpression promoted differentiation of human embryonal carcinoma cells [142]. Lastly, our findings are in apparent contrast with a recent work showing that *Alu* expression is downregulated in response to serum stimulation of human skin fibroblasts, establishing a negative relationship between *Alu* transcription and fibroblasts proliferation [58]. However, in that study no distinction was performed between Pol III-transcribed *Alu*s and *Alu* sequences embedded in Pol II transcripts. Therefore, *Alu* downregulation might have been indirectly caused by the inhibition of Pol II transcription, whereas our study unequivocally assesses the effects of *Alu* elements transcribed by Pol III.

Overall, the diverse effects in response to *Alu* overexpression could depend on the cell lineage and/or transformation state. Therefore, cell cycle induction by *Alu* transcripts might represent a new phenotype, since *Alu* overexpression has never been performed in primary human fibroblasts.

The observed changes in gene expression are most likely due to an effect *in trans* exerted by the overexpressed *Alu* sequences. Pol II inhibition, the modulation of the chromatin conformational state and the regulation of RNA stability are possible mechanisms exploited by *Alu*Sq2 and *Alu*Sx to induce differential gene expression. Especially, mRNA degradation might occur through a Staufen1 mRNA decay of genes downregulated by *Alu*Sq2, whereas a miRNA-like sponging mechanism might occur for *Alu*Sx upregulated genes. Indeed, we were able to verify the presence of targeting sites for two different miRNAs at the *AluSx* sequence. Interestingly, both miRNAs were found to potentially target the PRR11 gene, which is upregulated in our dataset. miRNAs annealing to *AluSx* might hinder their binding to PRR11, therefore inhibiting the formation of a silencing miRISC complex and the consequent mRNA degradation.
In conclusion, this work demonstrates that the oncoviral protein small E1A is able to enhance *Alu* transcription, possibly through e1a binding to the p400 protein. This interaction would stabilize p400 association to TIP60, which drives p400 on *Alu* elements that are marked by H3K4me1. The TIP60-p400 complex would finally lead to *Alu* activation.
e1a is known to induce extensive epigenetic changes in the host cell. *Alu* RNAs could play a role in this, as recent evidence supports an influence of SINE RNA in the spatial organization of chromatin.
Since the overexpression of *Alu* sequences leads to an increased cycling of human fibroblasts, it is tempting to speculate that Ad5 enhances *Alu* transcription to promote cell proliferation, exploiting endogenous SINEs for its own benefit. Indeed, the interaction of *Alu* RNAs with the host immune system is remarkably similar to what is observed for the Adenoviral Pol III VAI gene.
Even though our data support the idea that Ad5 might exploit the proliferative effects of Alu RNA, the maintenance of such a high copy number in the human genome might still be linked to *Alus'* important roles in cell processes. However, based on our results,

it seems more likely that *Alus* bring their positive contribution through *cis* effects on gene regulation networks, rather than *trans* mechanism exerted by *Alu* RNAs.

Supplementary information I

Alu RNA profiling in Adenovirus 5 infected cells

TableS1. *AluSp* and *AluSc* genomic coordinates and nucleotide sequences analysed through RT-qPCR. The genomic coordinates of the *Alu* sequences correspond to the annotated element. The 3′ trailer is in bold.

Name of *Alu*	*AluSp*
Genomic coordinates	chr5: 173764138-173764434
Genome assembly	GRCh38/hg38

***Alu* sequence:**

GGCCAGGCACAGTGGCTCACGTGTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGA
CAGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGACCAACATGGAGAAACCCCG
TCTCTACTAAAAATACAAAAATTAGCCGCGTTTGGTGGTGGGTGCCTGTAATCCCAG
CTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGACAGACATTGTGG
TAAGCCGAGATCATGCCACTGCACTCCAGCCTGGGCAACAAGAGTGAAAGTCCATC
TCCAAAAAAAAAAAAA**TCACTCTGTCGCTTCCCAGTTGCACACTCTATTCACTGGG**
GAGCCCTGCGTGAACAGAGCAGAGTTTTAAACACTGCTTTAAAAGCAGTGGCCA
AG

Name of *Alu*	*AluSc*
Genomic coordinates	chr18: 49460832-49461145
Genome assembly	GRCh38/hg38

***Alu* sequence:**

AGCCGTGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTAGGAGGCCGAGGTGGA
CGGATCACGAGGTCAGGAGATCGAGACCATCCTGGCCAGCATGGTGAAACCCCGTC
TCCACTAAAAATACAAAAAGTAGCTGTTTGAGACAGGATAATCGTTTGAACCCGAG
AGGCGGAGATTGCAGTGAGTCAAAAAGTAGCTGTCTGAGGCAGGAGAATCGTTTGA
ACCCGAGAGGCGGAGATTGCAGTGAGTCGAGATCGGGCCACTGCACTCCAGCCTTG
CGACAGAGCTCGACTCTTCTCAAAAAAAAAAAAA**GAGAGAGAGAGAGAAGCGCA**
AAAACCTCACCTGCCGGGTGCTAGAGTCAACCTCACCTGCCAGGTGCTAGAGTC
CTCGCTCATGCTACACCGCTCGTTCTGGCGGCTTAACCTTATTAAGGCCTC

Table S2. List of primers used for *Alu* and Pol III gene expression analysis in RT-qPCR.

Primer	Sequence (5′-3′)
*Alu*Sp_qRT_fw	TCACTCTGTCGCTTCCCAGTTG
*Alu*Sp_qRT_rev	TTAAAACTCTGCTCTGTTCACGCAG
*Alu*Sc_qRT_fw	GCTAGAGTCAACCTCACCTGC
*Alu*Sc_qRT_rev	TTAAGCCGCCAGAACGAGC
U1_fw	AGGGCGAGGCTTATCCATTG
U1_rev	GCAGTCGAGTTTCCCACATTTG
7SL_fw	CACCAGGTTGCCTAAGGAGG
7SL_rev	GCTGGAGTGCAGTGGCTATTC
U6_fw	CGCTTCGGCAGCACATATAC
U6_rev	AAAATATGGAACGCTTCACGA

Supplementary information II

Gene Expression analysis on fibroblasts overexpressing *Alu* sequences

TableS1. *AluSq2* and *AluSx* genomic coordinates and nucleotide sequences of the cloned elements into the lentivirus vector. The genomic coordinates of *Alu* sequences correspond to the annotated element. The 3′ trailer is in bold. In the Control sequence the portion deriving form *Alu*Sx is in bold.

Name of *Alu*	*AluSq2*
Genomic coordinates	chr1:61057625-61057914
Genome assembly	GRCh38/hg38

***Alu* sequence:**

GGCCAGGCGCTGTGGCTCACGCCTGTAATCCCAACACTTTGGGAGGCCGAGGCGAG
TGGATCACCTGAGGTCAGGAGTTCGCGACCAGCCTGACCCACATGGTGAAACCCCG
TCTCTACTAAAGTTAGCCAGACGTGGTGGCCGGCGCCTGTGATCTCAGCTACTCGGG
AGGCTGAGGCAGGAGAATCGCTTGTACCCGGGAGGCGAGGTTGCAGTGAGCCGAGA
TCGCGCCATTGCAGTCCAGCCTGGGCGACAAGAGCGAAACTCCGTCTAAAAAAAAA
AAAAAAAAA**GTGTCACCTCCCCATCTGCAAAGGTCTGGCCTCCTGAAAGCTCAG**
GAAACGGTGGGGCCATTTTT

Name of *Alu*	*AluSx*
Genomic coordinates	chr16:56419511-56419806
Genome assembly	GRCh38/hg38

***Alu* sequence:**

GGCCAGGCGTGGTGGCTCATGCCTGTAATCTCAGCACTTTGGGAGGCCAAGGTTAGT
GGATCACCTGAGGTCAGCAGTTCAAGACCAGCCTGGCCAACATGGTGAAATCCCGT
GTCTACTAAAAATACAAAAAATTAGCTGGGCATGGTGGTGCACACCTGTAATCTCAG
CTACTTGGGTGGCTGAGGCAGGAGAATTGCTTGAACCCAGGAGGCGGAGATTGCAG
TGAGCCGAGATTGTGCCATTGCACTCCAGCCTGGGCGACAGAGAAAGACTCTGACT
CAAAAAAAAAAAAAA**TTCAACTATATTAAAACACTTCAGAATGTTTCTCATAGCTA**
TAGTGCTGTCAAAGCTTCAAGTCATTTTTT

Control sequence:

GACCAGCGAATACCTGTTCCGTCATAGCGATAACGAGCTCCT
GCACTGGATGGTGGCGCTGGATGGTAAGCCGCTGGCAAGCGGTGAAGTGCCTCTGG
ATGTCGCTCCACAAGGTAAACAGTTGATTGAACTGCCTGAACTACCGCAGCCGGAG
AGCGCCGGGCAACTCTGGCTCACAGTACGCGTAGTGCAACCGAACGCGACCGCATG
GTCAGAAGCCGGGCACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAAAACCTCA
GTGTGACGCTCCCC**GCTTCAAGTCATTTTTT**

Table S2. List of primers used in the cloning strategy of *Alu* and control sequences into the lentiviral vector and used for the evaluation of overexpression of *Alu* and control sequences in Real Time PCR. The primer pairs are listed based on their use in the amplification of the sequence of interest from a genomic template, cloning into the pSUPER.GFP/neo vector or cloning into the lentiviral vector. The sequence for the *BglII*, *XhoI*, *HindIII*, *XbaI* or *BamHI* restriction enzymes is underlined. Nucleotide sequence is shown from the 5′ end to the 3′ end.

Primers for DNA amplification from the genomic template	
Primer	Sequence (5′-3′)
*Alu*Sq2_chr1_fw	GCCCCAGGTGATCTCTACC
*Alu*Sq2_chr1_rev	GTCCTCGGAGCCGCTAATTT
*Alu*Sx_Fw	CCCTTAACTTTTGTACCCTGAGC
*Alu*Sx_Re	CACTCTGAACGGGGACAAGTA
LacZ_fw	TAAATATAAAAGATCTGACCAGCGAATACCTGTTCC
LacZ_re2	TTTTTTTTGGGGAGCGTCACACTGAG

Primers for cloning into the pSUPER.GFP/neo vector	
Primer	Sequence (5′-3′)
*Alu*Sq2_pSUPER_Bgl_fw	TAAATATAAAAGATCTGGCCAGGCGCTGTGGCT
*Alu*Sq2_pSUPER_Xho_rev	AATTATTTTACTCGAGAAAAATGGCCACCACCGTTTCC
*Alu*Sx_Fw2	TAAATATAAAAGATCTGGCCAGGCGTGGTGG
*Alu*Sx_Re2	AATTATTTTACTCGAGAAAAAATGACTTGAAGCTTTGACAGCA
LacZ_fw	TAAATATAAAAGATCTGACCAGCGAATACCTGTTCC
LacZ_re3	AATTATTTTAAAGCTTTTTTTTTTTTTTTTTTGGGGAGCGTCAC
Primers for subcloning into the lentiviral vector	
Primer	Sequence (5′-3′)
5′_XbaI_H1_prom	GAGTTCTAGAGAACGCTGACGTCATCAACCC
3′_BamHI_*Alu*Sq2	CCTCCGGATCCAAAAATGGCCCCACCGTTTCC
3′_BamHI_*Alu*Sx	CCTCCGGATCCAAAAAATGACTTGAAGCTTTG
3′_BamHI_control	GAGTGGATCCAAAAAATGACTTGAAGC
3′_BamHI_no_insert	CCTCCGGATCCGTCGACGGTATCGATAAGCTTAG

Continued on next page

Continues from previous page

Primers used in RT-qPCR	
Primer	Sequence (5′-3′)
*Alu*Sq2_qRT_fw	AAGTGTCACCTCCCCATCTG
*Alu*Sq2_qRT _rev	ACCACCGTTTCCTGAGCTT
*Alu*Sx_qRT_fw	AATTCAACTATATTAAAACACTTCAGA
*Alu*Sx_qRT_rev	GACTTGAAGCTTTGACAGCA
U1_fw	AGGGCGAGGCTTATCCATTG
U1_rev	GCAGTCGAGTTTCCCACATTTG

Table S3. List of Differentially Expressed genes (|log_2 Fold Change>0.5| and *p*-adjusted value < 0.001) detected in IMR90 cells overexpressing *AluSq2* sequence. Genes are sorted based on log_2 FC values in ascending order.

Gene symbol	log2 Fold Change	p-adjusted value	Gene symbol	log2 Fold Change	p-adjusted value	Gene symbol	log2 Fold Change	p-adjusted value
FAM198B	-1.123	6.92E-16	SACS	-0.569	3.72E-07	NUF2	0.565	2.83E-06
SCN7A	-1.089	1.24E-12	KCNMA1	-0.568	1.44E-06	CENPA	0.566	6.43E-05
NCR3LG1	-1.031	1.30E-12	TIMP3	-0.566	1.25E-11	SOCS2	0.567	1.44E-07
SLC14A1	-1.022	6.07E-13	ZNF853	-0.564	6.38E-04	AREG	0.569	9.62E-04
ASS1	-0.987	9.75E-12	PDP2	-0.564	2.48E-05	CEP55	0.575	2.77E-09
COL15A1	-0.979	7.58E-27	IRS2	-0.564	4.26E-05	FAM167A	0.583	6.34E-07
UNC5B	-0.880	2.18E-14	TGFBR3	-0.557	1.69E-04	AMIGO2	0.583	2.26E-06
CREB5	-0.852	3.15E-08	KIAA1217	-0.556	4.80E-05	SHC4	0.584	2.36E-06
POSTN	-0.847	6.02E-08	IL6ST	-0.554	5.70E-04	PDE3A	0.589	5.55E-04
SCD	-0.842	3.34E-21	CPA4	-0.548	5.44E-06	COLEC10	0.593	7.42E-05
IL16	-0.835	9.58E-09	NRP1	-0.544	9.25E-11	CDKN3	0.597	4.82E-06
VCAM1	-0.823	6.34E-08	RAMP1	-0.541	2.39E-05	CSGALNACT1	0.599	7.01E-08
ADAMTS8	-0.799	3.15E-08	ADM2	-0.540	1.31E-04	EDN1	0.601	3.53E-04
F2RL2	-0.797	1.06E-10	TMEM130	-0.539	4.84E-04	NRGN	0.601	5.86E-06
SH2D5	-0.795	1.51E-09	TNRC6C	-0.537	8.59E-05	SERPINI1	0.611	3.14E-04
USP53	-0.776	1.11E-10	ARNT2	-0.535	8.12E-08	PMEL	0.623	2.15E-04
HDAC9	-0.775	1.30E-06	BACH1	-0.530	6.19E-06	MYOCD	0.624	1.49E-07
B3GALT2	-0.762	3.41E-07	COL6A1	-0.530	1.39E-12	STC1	0.627	3.44E-11
MASP1	-0.758	6.33E-09	BAALC	-0.527	3.56E-05	PTTG1	0.628	1.44E-13
ROR1	-0.753	1.78E-08	ATP8B1	-0.525	6.38E-06	FRMD3	0.642	1.12E-04
ALPL	-0.748	4.26E-06	HIPK2	-0.522	1.28E-06	HHIP	0.648	5.38E-05
WISP1	-0.740	5.65E-12	ALDH1L2	-0.520	2.51E-08	GREM2	0.650	4.77E-06
CASS4	-0.737	8.54E-11	UAP1L1	-0.520	2.83E-06	TMSB15A	0.653	9.36E-06
SVEP1	-0.733	5.10E-14	SEPT6	-0.519	1.41E-04	CDK2	0.666	4.48E-16
CRYBG1	-0.731	1.18E-07	ICAM1	-0.516	2.20E-05	CCND3	0.667	5.88E-19
TMEM119	-0.730	5.86E-06	LRIG1	-0.515	1.38E-05	CDCA3	0.681	1.15E-11
CCL2	-0.716	7.67E-22	VPS13C	-0.514	2.87E-05	SOCS2-AS1	0.689	1.75E-06
CLDN11	-0.715	1.50E-08	ANKRD33B	-0.511	4.35E-04	IL11	0.722	1.61E-15
RUNX1T1	-0.710	2.40E-06	SLFN5	-0.509	1.87E-08	FAM105A	0.724	4.06E-08
NFAT5	-0.703	9.09E-09	ZNF106	-0.508	1.32E-05	ENO3	0.733	5.71E-08
CORIN	-0.702	2.86E-11	RELB	-0.507	1.44E-04	AL645608.1	0.735	3.52E-06
CLCN5	-0.693	2.38E-06	VAT1L	-0.506	1.06E-04	MANF	0.735	2.75E-19
TENM4	-0.691	9.03E-07	DCHS1	-0.501	3.82E-07	ANKRD1	0.748	7.33E-07
LSAMP	-0.688	1.47E-07	ITGBL1	-0.501	5.02E-06	ID3	0.771	1.35E-06
ELOVL2	-0.673	1.76E-05	GLA	0.504	1.19E-06	DNER	0.778	3.20E-07
FNIP1	-0.663	5.95E-08	PRRX2	0.505	1.23E-07	P4HA3	0.783	4.07E-09
SMAD1	-0.653	6.44E-06	A4GALT	0.505	1.88E-04	IL24	0.787	1.06E-06
RGS17	-0.651	3.64E-05	TK1	0.511	3.82E-10	CYP2S1	0.810	8.70E-10
TRPA1	-0.647	4.54E-11	CDK1	0.513	1.94E-06	SAMD11	0.847	2.96E-09
KIAA1644	-0.645	9.99E-05	LYPD1	0.516	3.31E-07	AL353653.1	0.848	6.63E-09
FBN2	-0.645	1.62E-11	RIPOR3	0.517	3.27E-06	RGS5	0.860	5.52E-09
NAV2	-0.643	3.66E-07	CAMK1	0.519	1.47E-06	CRYAB	0.881	2.80E-12
PSAT1	-0.638	9.32E-14	NDRG1	0.519	3.76E-08	MMP1	0.897	4.52E-11
BMP2K	-0.636	4.75E-08	SGK1	0.523	6.38E-07	ATOH8	0.899	3.54E-12
CACNA1H	-0.628	1.09E-04	MXD3	0.525	5.51E-04	SCG2	0.917	4.72E-16
TRIB3	-0.627	6.79E-10	UBE2T	0.526	1.57E-05	TGM2	0.918	4.56E-43
DHCR24	-0.624	1.56E-09	NOX4	0.527	8.22E-04	TM4SF1	0.972	6.66E-37
ITGA11	-0.624	5.57E-15	CCNA2	0.529	1.22E-07	ANGPTL4	1.030	2.46E-30
HGF	-0.621	2.36E-13	CABLES1	0.530	1.26E-06	CRLF1	1.030	9.75E-12
LURAP1L	-0.621	1.28E-04	KIF20A	0.530	4.24E-09	SLCO2A1	1.041	9.38E-13
C7	-0.620	4.77E-06	CDC20	0.530	6.74E-10	FGL2	1.046	3.99E-19
IL32	-0.618	7.71E-05	SPAG5	0.535	1.66E-09	NR4A1	1.093	5.03E-22
FLCN	-0.602	5.68E-08	CENPW	0.537	3.84E-05	TFPI2	1.303	7.19E-20
INSIG1	-0.596	5.25E-06	GPRC5A	0.538	1.78E-05	MMP3	1.321	2.45E-25
SLIT2	-0.595	1.76E-05	PRC1	0.540	7.74E-12			
TNC	-0.595	3.81E-12	GNG11	0.541	5.19E-09			
GPC6	-0.589	1.91E-06	UBE2C	0.544	6.14E-08			
AC012513.3	-0.585	5.11E-04	PTPRN	0.546	1.57E-05			
MSMO1	-0.579	3.82E-06	SCNN1D	0.546	2.93E-04			
LIFR	-0.578	7.46E-07	CCNB2	0.548	4.97E-08			
PAG1	-0.576	2.37E-09	DLGAP5	0.550	2.70E-09			
CDCP1	-0.575	4.38E-07	SERPINE2	0.552	2.82E-08			
KCNK2	-0.575	2.99E-04	PSMC3IP	0.552	3.01E-05			
SLC4A7	-0.571	6.24E-06	SAPCD2	0.553	1.53E-04			
JMJD1C	-0.570	3.31E-07	PIMREG	0.553	1.75E-07			
COL3A1	-0.570	1.51E-09	KYNU	0.561	8.13E-04			
SRCAP	-0.570	8.21E-04	A2M	0.561	2.70E-09			

Table S4. List of Differentially Expressed genes (|log2 FC>0.5| and *p*-adjusted value < 0.001) detected in IMR90 cells overexpressing *AluSx* sequence. Genes are sorted based on log_2 FC values in ascending order.

Gene symbol	log2 Fold Change	p-adjusted value	Gene symbol	log2 Fold Change	p-adjusted value	Gene symbol	log2 Fold Change	p-adjusted value
CCL2	-1.498	2.50E-97	C19orf73	-0.573	1.19E-09	TIMELESS	0.542	2.12E-08
FAM198B	-1.251	6.62E-20	EPAS1	-0.572	2.73E-14	TRIP13	0.544	3.30E-07
ADAMTS8	-1.251	6.20E-20	MAN1C1	-0.572	1.24E-07	PMEPA1	0.545	6.58E-06
LIF	-1.038	4.58E-17	RPS6KA2	-0.569	7.66E-08	CENPI	0.548	2.00E-04
MASP1	-1.036	1.24E-16	ADAMTS14	-0.567	8.89E-04	NCAPD2	0.549	1.38E-10
ASS1	-1.013	1.37E-12	C5orf66	-0.562	3.58E-04	PODXL	0.549	3.64E-04
CREB5	-0.979	6.10E-11	HGF	-0.560	5.32E-11	DNER	0.551	7.20E-04
COL15A1	-0.924	5.32E-24	PDE7B	-0.557	6.13E-06	PRR11	0.554	3.96E-06
SCD	-0.921	1.73E-25	IL32	-0.557	4.52E-04	APCDD1L	0.555	8.20E-04
IL16	-0.912	1.39E-10	DHCR24	-0.556	8.96E-08	DIAPH3	0.556	2.36E-07
TMEM130	-0.904	1.07E-10	IFIT1	-0.556	4.08E-04	ATOH8	0.558	6.37E-05
CASS4	-0.890	6.95E-16	B3GALT2	-0.553	4.34E-04	DBF4B	0.558	7.27E-05
ALPL	-0.883	1.91E-08	MYC	-0.550	1.65E-05	TM4SF1	0.561	3.47E-12
VCAM1	-0.879	3.86E-09	TNIK	-0.549	1.97E-04	AC145098.2	0.564	9.04E-04
SHROOM2	-0.874	3.07E-15	LRIG1	-0.535	4.29E-06	PRRX2	0.567	8.50E-10
CCDC102B	-0.866	1.12E-13	FLCN	-0.534	1.51E-06	ESAM	0.568	1.48E-04
SVEP1	-0.834	2.55E-18	FAM49A	-0.533	5.67E-04	CYR61	0.569	2.30E-10
SCN7A	-0.830	1.48E-07	RDH10	-0.531	3.29E-05	RECQL4	0.569	2.07E-06
NNMT	-0.813	1.56E-15	PDGFRL	-0.531	2.14E-06	UBE2S	0.569	1.62E-06
NCR3LG1	-0.788	1.12E-07	EHD3	-0.531	9.34E-09	RRM2	0.570	3.30E-09
KCNJ2	-0.771	1.32E-06	CRYBG1	-0.529	2.55E-04	BUB1	0.571	1.52E-07
CLDN11	-0.765	6.17E-10	MCC	-0.526	3.94E-07	SMTN	0.572	2.96E-10
SH2D5	-0.763	4.92E-09	MAMLD1	-0.524	8.34E-05	BIRC5	0.574	8.94E-11
RAMP1	-0.758	2.61E-10	BMP2K	-0.524	9.93E-06	P4HA3	0.575	4.02E-05
F2RL2	-0.747	1.19E-09	SHC3	-0.524	8.07E-08	RACGAP1	0.575	1.07E-10
SMAD1	-0.745	9.82E-08	MT1E	-0.522	2.40E-07	RAD54L	0.577	6.85E-05
TMEM119	-0.737	3.75E-06	ARNT2	-0.521	1.29E-07	PKMYT1	0.578	4.83E-06
KIAA1644	-0.732	5.20E-06	ANGPTL2	-0.518	2.41E-08	KIF4A	0.578	1.90E-08
SOCS1	-0.729	1.44E-07	SHISA3	-0.515	4.34E-04	FAM214B	0.580	2.91E-10
CORIN	-0.720	3.81E-12	PRR16	-0.514	1.14E-06	ZWINT	0.581	2.50E-10
WISP1	-0.718	1.39E-11	TENM4	-0.511	5.27E-04	NCAPG	0.583	3.41E-08
FIBIN	-0.716	3.80E-09	PLPP3	-0.507	1.04E-07	AL645608.1	0.583	4.42E-04
LSAMP	-0.696	6.20E-08	UBA7	-0.507	1.59E-05	CDKN2D	0.583	1.35E-05
ICAM1	-0.695	1.24E-09	DAB2	-0.504	1.44E-10	ULBP2	0.586	1.54E-04
ROR1	-0.688	2.40E-07	USP53	-0.503	8.72E-05	CENPW	0.586	3.47E-06
TGFBR3	-0.669	2.20E-06	CPA4	-0.502	3.46E-05	SHCBP1	0.587	1.09E-08
GDF5	-0.651	9.38E-05	DAAM2	-0.502	3.88E-04	MMP10	0.588	2.67E-04
ITGBL1	-0.648	5.14E-10	LRP11	0.501	3.27E-09	LYPD1	0.589	2.02E-09
LITAF	-0.639	1.17E-14	INA	0.501	3.60E-04	MMP1	0.597	4.02E-05
ADORA2B	-0.639	4.82E-05	HERC2P2	0.503	7.25E-04	RNASEH2A	0.597	2.18E-08
HGNC:18790	-0.639	1.37E-04	TONSL	0.505	1.07E-04	ENO3	0.599	1.59E-05
FNIP1	-0.635	1.65E-07	CDK2	0.505	1.96E-09	DLGAP5	0.599	2.95E-11
SEPT6	-0.634	1.00E-06	NCAPG2	0.507	4.84E-08	ITGB3	0.599	3.42E-07
THBS2	-0.633	6.36E-06	FKBP4	0.507	1.36E-08	CCNB2	0.599	8.57E-10
GAS1	-0.630	1.56E-04	KIF22	0.508	1.30E-07	IL21R	0.600	2.36E-04
C7	-0.624	2.88E-06	NPC1	0.509	1.59E-05	PSMC3IP	0.601	2.93E-06
CEMIP	-0.623	4.30E-05	SPRY4	0.510	9.45E-07	CENPH	0.606	3.25E-05
PALM	-0.621	2.06E-04	STC1	0.511	1.29E-07	SUN2	0.606	4.67E-14
INSIG1	-0.618	1.52E-06	SNCAIP	0.512	5.64E-05	SKA3	0.606	2.40E-07
SQOR	-0.606	3.44E-05	MCM5	0.515	1.30E-08	HMMR	0.608	9.54E-07
SLIT2	-0.605	9.61E-06	RAD51	0.518	6.27E-05	TK1	0.611	7.67E-15
PCDHGC3	-0.603	1.21E-08	KIAA1524	0.520	1.64E-05	CENPK	0.614	3.26E-05
FABP3	-0.600	1.09E-04	FANCG	0.522	1.56E-05	GNB3	0.616	1.60E-05
GPAT3	-0.599	1.17E-04	ABCB1	0.523	5.68E-04	ANKRD1	0.619	7.27E-05
MSMO1	-0.591	1.63E-06	CAMK1	0.524	8.22E-07	CRIP2	0.619	4.13E-14
SQSTM1	-0.591	1.01E-14	E2F1	0.527	3.82E-06	TMSB15A	0.624	2.40E-05
LIFR	-0.590	2.40E-07	PITPNM3	0.528	9.93E-07	CCNB1	0.625	5.82E-14
DDIT4L	-0.590	4.55E-04	SERPINE2	0.529	9.25E-08	KIF18B	0.626	1.44E-07
AL355075.4	-0.589	5.52E-04	WDR62	0.530	2.21E-05	PBK	0.628	3.89E-08
HDAC9	-0.589	4.78E-04	HAPLN3	0.530	5.09E-05	CDK1	0.628	1.05E-09
IGFBP2	-0.588	3.50E-04	SCNN1D	0.532	4.24E-04	SAPCD2	0.629	7.47E-06
IRS2	-0.583	1.65E-05	PTPRN	0.532	2.33E-05	KIF2C	0.634	1.05E-11
PLXNC1	-0.582	4.22E-04	ANLN	0.533	1.32E-06	CHTF18	0.635	8.30E-07
ANKRD33B	-0.581	3.27E-05	GTSE1	0.533	2.07E-06	LCE2A	0.636	1.05E-04
TRIB3	-0.578	1.17E-08	MXD3	0.534	3.77E-04	TACC3	0.636	4.69E-15
GNPDA1	-0.578	5.34E-07	CDC25C	0.535	3.98E-04	PIMREG	0.640	3.15E-10
HTR2B	-0.575	4.08E-04	BUB1B	0.537	1.36E-07	SAMD11	0.642	1.61E-05
SFRP1	-0.574	1.03E-10	RHOB	0.542	8.84E-11	MANF	0.643	6.53E-15

Continued on next page

Table S4 - Continues from previous

Gene symbol	log2 Fold Change	p-adjusted value	Gene symbol	log2 Fold Change	p-adjusted value
POC1A	0.644	3.59E-08	PDGFB	0.717	1.06E-05
COL11A1	0.647	7.94E-13	CRYAB	0.725	1.70E-08
TPX2	0.648	5.70E-18	NOX4	0.726	5.20E-07
UBE2T	0.651	1.72E-08	EDN1	0.732	4.83E-06
PLK1	0.652	3.59E-11	SPDL1	0.738	2.76E-16
ELN	0.653	8.70E-05	PTTG1	0.741	2.63E-19
DEPDC1	0.664	1.93E-07	TFPI2	0.744	1.25E-06
SPAG5	0.667	3.96E-15	SCG2	0.744	9.53E-11
FRMD3	0.668	5.08E-05	TROAP	0.745	1.57E-12
PRDM1	0.669	3.32E-07	KLF2	0.746	4.33E-08
GREM2	0.670	1.63E-06	PRC1	0.752	2.02E-23
KIF20A	0.672	6.92E-15	CEP55	0.754	2.76E-16
CENPA	0.673	6.22E-07	PRR5L	0.775	1.27E-06
NUF2	0.674	4.89E-09	CDCA3	0.779	1.46E-15
UBE2C	0.676	1.71E-12	ADAM19	0.783	2.94E-21
MYBL2	0.678	9.76E-14	MYOCD	0.785	4.06E-12
CHL1	0.679	1.59E-08	FAM105A	0.824	1.04E-10
CDKN3	0.684	5.15E-08	IQGAP3	0.832	9.96E-17
CCNA2	0.689	3.45E-13	CYP2S1	0.853	4.34E-11
NRGN	0.690	5.99E-08	AMIGO2	0.900	3.96E-15
CDH6	0.709	6.82E-15	SLCO2A1	0.944	1.07E-10
CDH2	0.710	1.49E-14	ANGPTL4	1.143	7.48E-38
CDC20	0.711	3.47E-18	CRLF1	1.143	1.05E-14
FOXM1	0.714	3.34E-15	MMP3	1.363	3.94E-27

Table S5. Table A4. List of Differentially Expressed genes (|log2FC>0.5| and *p*-adjusted value < 0.001) detected in IMR90 cells overexpressing a control sequence. Genes are sorted based on log_2 FC values in ascending order.

Gene symbol	log2 Fold Change	p-adjusted value	Gene symbol	log2 Fold Change	p-adjusted value	Gene symbol	log2 Fold Change	p-adjusted value
ASS1	-1.045	4.43E-13	FNIP1	-0.534	3.40E-05	LCE2A	0.611	3.24E-04
FAM198B	-1.029	2.51E-13	C7	-0.534	1.68E-04	PTPRN	0.613	9.63E-07
CORIN	-0.944	3.02E-20	BAALC	-0.526	5.53E-05	CRIP2	0.614	1.78E-13
HDAC9	-0.915	5.79E-09	HMGCS1	-0.523	5.88E-07	ITGA8	0.617	1.37E-06
POSTN	-0.885	1.58E-08	ITGBL1	-0.523	2.10E-06	PRDM1	0.631	3.26E-06
SCN7A	-0.856	8.30E-08	TIMP3	-0.508	3.31E-09	HES4	0.632	3.61E-06
SCD	-0.849	1.38E-21	NRP1	-0.506	3.54E-09	EDN1	0.634	1.87E-04
COL15A1	-0.827	3.77E-19	LRRC17	-0.500	1.28E-09	CDK2	0.641	6.26E-15
B3GALT2	-0.823	3.16E-08	MYOCD	0.500	7.79E-05	MEG9	0.642	1.14E-05
USP53	-0.806	2.37E-11	PMEPA1	0.504	7.51E-05	CRYAB	0.643	1.60E-06
SLC14A1	-0.806	3.96E-08	CDK2AP2	0.507	3.71E-05	QRICH2	0.651	1.30E-04
CREB5	-0.790	4.93E-07	HDAC5	0.509	2.65E-06	GNB3	0.652	6.50E-06
RAMP1	-0.783	1.94E-10	CHL1	0.510	1.00E-04	CCND3	0.658	1.99E-18
MSMO1	-0.772	1.98E-10	MFAP2	0.511	9.63E-05	NRGN	0.664	4.47E-07
OXTR	-0.767	2.62E-08	LINC00475	0.515	4.67E-04	MANF	0.666	8.25E-16
CCL2	-0.767	3.98E-25	MICAL1	0.515	8.51E-10	ENO3	0.680	8.91E-07
KCNK2	-0.767	4.79E-07	CSGALNACT1	0.518	7.68E-06	IL11	0.685	6.88E-14
F2RL2	-0.743	3.54E-09	NFKBIZ	0.520	9.63E-05	DNER	0.686	1.41E-05
MASP1	-0.739	2.16E-08	ZNF692	0.523	4.97E-05	MMP10	0.705	7.68E-06
INSIG1	-0.736	8.16E-09	ACVRL1	0.524	8.23E-06	ID3	0.718	1.18E-05
SH2D5	-0.732	4.12E-08	A4GALT	0.524	1.25E-04	FAM105A	0.734	2.99E-08
CRYBG1	-0.720	2.44E-07	SYNM	0.526	4.29E-07	SCNN1D	0.735	2.86E-07
SLIT2	-0.677	8.39E-07	CABLES1	0.534	1.27E-06	BMP2	0.738	4.30E-10
ELOVL2	-0.669	3.00E-05	FAM167A	0.535	8.95E-06	KYNU	0.754	2.29E-06
CASS4	-0.659	1.35E-08	CORO6	0.537	3.14E-04	AL645608.1	0.762	1.74E-06
VCAM1	-0.646	6.72E-05	SLC26A6	0.538	2.86E-07	TM4SF1	0.778	2.32E-23
DHCR24	-0.645	4.97E-10	LY6K	0.541	8.23E-05	SAMD11	0.791	5.32E-08
PSAT1	-0.638	9.17E-14	MAMDC4	0.543	8.06E-04	P4HA3	0.791	3.78E-09
IL16	-0.637	3.94E-05	KLF2	0.552	2.19E-04	ATOH8	0.803	1.38E-09
NCR3LG1	-0.628	8.30E-05	NDRG1	0.556	3.62E-09	STC1	0.806	1.44E-18
LSAMP	-0.627	2.98E-06	KLHL17	0.558	1.81E-06	COLEC10	0.811	1.24E-08
BMP2K	-0.619	1.49E-07	RGS18	0.561	6.14E-04	SCG2	0.813	1.63E-12
CLDN11	-0.612	2.86E-06	SOCS2-AS1	0.565	2.29E-04	A2M	0.824	4.00E-20
ROR1	-0.602	1.78E-05	ANO8	0.570	1.91E-06	FGL2	0.825	1.18E-11
IL32	-0.601	1.83E-04	CDCA3	0.572	3.68E-08	RIPOR3	0.846	2.50E-16
TRIB3	-0.601	5.30E-09	PRRX2	0.577	8.28E-10	MMP1	0.862	4.56E-10
SULF1	-0.589	4.90E-06	SAPCD2	0.578	9.09E-05	RGS5	0.867	5.30E-09
HTR2B	-0.581	4.32E-04	SERPINE2	0.580	5.30E-09	IL24	0.898	1.58E-08
TLR6	-0.577	7.24E-04	PMEL	0.581	9.49E-04	AL353653.1	0.922	2.38E-10
RGS17	-0.569	6.15E-04	PITPNM1	0.593	4.30E-10	SLCO2A1	1.010	6.30E-12
NAV2	-0.569	1.41E-05	IQGAP3	0.594	2.45E-08	TGM2	1.014	3.41E-53
THBS2	-0.563	1.46E-04	AREG	0.599	6.02E-04	ANGPTL4	1.043	3.55E-31
ATP8B1	-0.551	2.44E-06	SHC4	0.600	1.37E-06	CYP2S1	1.060	5.70E-17
SYNPO2	-0.542	6.46E-04	GREM2	0.601	4.41E-05	TFPI2	1.195	1.22E-16
CPA4	-0.540	1.08E-05	JUNB	0.602	1.56E-08	CRLF1	1.197	6.66E-16
FABP3	-0.538	9.27E-04	CYGB	0.604	3.18E-07	NR4A1	1.256	1.61E-29
PRSS12	-0.536	4.74E-07	SGK1	0.606	4.31E-09	MMP3	1.486	2.61E-32

Table S6. List of DE genes with at least one *Alu* element embedded in their 3′ UTR.

Gene symbol	Number of Alu sequences	Number of sense Alu sequences	Number of anti-sense Alu sequences	% of anti-sense Alu sequences	log2FoldChange of DEG in AluSq2-overexpressing cells	AluSq2 padj	log2FoldChange of DEG in AluSx-overexpressing cells	AluSx padj
AluSq2_up								
GPRC5A	4	0	4	100.0	0.54	1.78E-05	0.09	7.17E-01
NOX4	1	1	0	0.0	0.53	8.22E-04	0.73	5.20E-07
IL11	1	0	1	100.0	0.72	1.61E-15	0.21	8.50E-02
ENO3	1	0	1	100.0	0.73	5.71E-08	0.60	1.59E-05
KYNU	3	0	3	100.0	0.56	8.13E-04	0.23	3.17E-01
SGK1	2	0	2	100.0	0.52	6.38E-07	0.28	2.07E-02
FGL2	1	1	0	0.0	1.05	3.99E-19	0.39	6.33E-03
HHIP	1	0	1	100.0	0.65	5.38E-05	0.30	1.42E-01
SAPCD2	1	1	0	0.0	0.55	1.53E-04	0.63	7.47E-06
AluSq2_dn								
IL32	1	0	1	100.0	-0.62	7.71E-05	-0.56	4.52E-04
CLDN11	1	0	1	100.0	-0.71	1.50E-08	-0.77	6.17E-10
TNC	1	1	0	0.0	-0.59	3.81E-12	-0.30	2.29E-03
HDAC9	1	0	1	100.0	-0.78	1.30E-06	-0.59	4.78E-04
MSMO1	1	0	1	100.0	-0.58	3.82E-06	-0.59	1.63E-06
PAG1	1	1	0	0.0	-0.58	2.37E-09	-0.24	4.06E-02
SRCAP	2	1	1	50.0	-0.57	8.21E-04	-0.26	2.49E-01
ATP8B1	2	2	0	0.0	-0.53	6.38E-06	-0.32	1.50E-02
CASS4	3	1	2	66.7	-0.74	8.54E-11	-0.89	6.95E-16
ICAM1	1	0	1	100.0	-0.52	2.20E-05	-0.70	1.24E-09
RGS17	1	1	0	0.0	-0.65	3.64E-05	-0.45	8.63E-03
NRP1	1	1	0	0.0	-0.54	9.25E-11	-0.32	4.70E-04
NFAT5	4	3	1	25.0	-0.70	9.09E-09	-0.36	1.23E-02
ZNF106	2	2	0	0.0	-0.51	1.32E-05	-0.40	1.01E-03
CRYBG1	2	1	1	50.0	-0.73	1.18E-07	-0.53	2.55E-04
C7	1	0	1	100.0	-0.62	4.77E-06	-0.62	2.88E-06
ADM2	2	2	0	0.0	-0.54	1.31E-04	-0.22	2.49E-01
CPA4	1	0	1	100.0	-0.55	5.44E-06	-0.50	3.46E-05
VPS13C	1	0	1	100.0	-0.51	2.87E-05	-0.35	1.08E-02
ALDH1L2	1	0	1	100.0	-0.52	2.51E-08	-0.29	6.99E-03
ITGA11	3	1	2	66.7	-0.62	5.57E-15	-0.32	3.58E-04
SLC14A1	2	2	0	0.0	-1.02	6.07E-13	-0.38	2.89E-02
USP53	2	1	1	50.0	-0.78	1.11E-10	-0.50	8.72E-05
CREB5	1	0	1	100.0	-0.85	3.15E-08	-0.98	6.10E-11
FLCN	1	0	1	100.0	-0.60	5.68E-08	-0.53	1.51E-06
CDCP1	1	0	1	100.0	-0.57	4.38E-07	-0.35	4.82E-03
F2RL2	2	0	2	100.0	-0.80	1.06E-10	-0.75	1.19E-09
TMEM130	1	1	0	0.0	-0.54	4.84E-04	-0.90	1.07E-10
SLFN5	3	3	0	0.0	-0.51	1.87E-08	-0.49	5.99E-08
SMAD1	1	1	0	0.0	-0.65	6.44E-06	-0.75	9.82E-08
CLCN5	1	0	1	100.0	-0.69	2.38E-06	-0.49	2.03E-03
ARNT2	1	1	0	0.0	-0.53	8.12E-08	-0.52	1.29E-07
PDP2	6	3	3	50.0	-0.56	2.48E-05	-0.30	6.65E-02

Continued next page

From previous page

TMEM119	1	0	1	100.0	-0.73	5.86E-06	-0.74	3.75E-06
IRS2	1	1	0	0.0	-0.56	4.26E-05	-0.58	1.65E-05
UAP1L1	1	0	1	100.0	-0.52	2.83E-06	-0.47	2.22E-05
FNIP1	1	0	1	100.0	-0.66	5.95E-08	-0.63	1.65E-07
AluSx_up								
FKBP4	1	1	0	0.0	0.47	2.26E-07	0.51	1.36E-08
NCAPD2	1	1	0	0.0	0.31	1.48E-03	0.55	1.38E-10
TACC3	1	1	0	0.0	0.49	7.35E-09	0.64	4.69E-15
ELN	1	1	0	0.0	0.50	5.17E-03	0.65	8.70E-05
RAD51	1	1	0	0.0	0.38	6.65E-03	0.52	6.27E-05
PRR11	3	3	0	0.0	0.21	1.93E-01	0.55	3.96E-06
GTSE1	2	2	0	0.0	0.27	5.18E-02	0.53	2.07E-06
WDR62	2	0	2	100.0	0.33	2.04E-02	0.53	2.21E-05
NOX4	1	1	0	0.0	0.53	8.22E-04	0.73	5.20E-07
BIRC5	1	1	0	0.0	0.42	7.56E-06	0.57	8.94E-11
IL21R	2	1	1	50.0	0.40	2.50E-02	0.60	2.36E-04
UBE2S	1	1	0	0.0	0.39	2.85E-03	0.57	1.62E-06
ENO3	1	0	1	100.0	0.73	5.71E-08	0.60	1.59E-05
TIMELESS	1	1	0	0.0	0.40	9.79E-05	0.54	2.12E-08
CDH6	1	0	1	100.0	0.38	2.15E-04	0.71	6.82E-15
TROAP	1	1	0	0.0	0.49	2.12E-05	0.75	1.57E-12
NPC1	2	2	0	0.0	0.40	1.24E-03	0.51	1.59E-05
NCAPG2	1	0	1	100.0	0.32	2.16E-03	0.51	4.84E-08
CENPH	1	0	1	100.0	0.48	2.04E-03	0.61	3.25E-05
DBF4B	1	1	0	0.0	0.41	8.07E-03	0.56	7.27E-05
PLK1	1	1	0	0.0	0.44	3.13E-05	0.65	3.59E-11
SHCBP1	1	1	0	0.0	0.35	2.56E-03	0.59	1.09E-08
RRM2	1	0	1	100.0	0.44	1.60E-05	0.57	3.30E-09
KIF18B	2	0	2	100.0	0.40	2.91E-03	0.63	1.44E-07
SAPCD2	1	1	0	0.0	0.55	1.53E-04	0.63	7.47E-06
AluSx_dn								
IL32	1	0	1	100.0	-0.62	7.71E-05	-0.56	4.52E-04
CLDN11	1	0	1	100.0	-0.71	1.50E-08	-0.77	6.17E-10
HDAC9	1	0	1	100.0	-0.78	1.30E-06	-0.59	4.78E-04
MSMO1	1	0	1	100.0	-0.58	3.82E-06	-0.59	1.63E-06
CASS4	3	1	2	66.7	-0.74	8.54E-11	-0.89	6.95E-16
ICAM1	1	0	1	100.0	-0.52	2.20E-05	-0.70	1.24E-09
PALM	2	2	0	0.0	-0.39	4.18E-02	-0.62	2.06E-04
CRYBG1	2	1	1	50.0	-0.73	1.18E-07	-0.53	2.55E-04
C7	1	0	1	100.0	-0.62	4.77E-06	-0.62	2.88E-06
IGFBP2	2	0	2	100.0	-0.31	1.17E-01	-0.59	3.50E-04
MAN1C1	1	0	1	100.0	-0.33	8.46E-03	-0.57	1.24E-07
RDH10	1	0	1	100.0	-0.26	8.87E-02	-0.53	3.29E-05
CPA4	1	0	1	100.0	-0.55	5.44E-06	-0.50	3.46E-05
USP53	2	1	1	50.0	-0.78	1.11E-10	-0.50	8.72E-05

Continued om next page

From previous page

CREB5	1	0	1	100.0	-0.85	3.15E-08	-0.98	6.10E-11
CCDC102B	2	1	1	50.0	-0.41	2.33E-03	-0.87	1.12E-13
FLCN	1	0	1	100.0	-0.60	5.68E-08	-0.53	1.51E-06
SQSTM1	1	0	1	100.0	-0.42	1.49E-07	-0.59	1.01E-14
F2RL2	2	0	2	100.0	-0.80	1.06E-10	-0.75	1.19E-09
TMEM130	1	1	0	0.0	-0.54	4.84E-04	-0.90	1.07E-10
SMAD1	1	1	0	0.0	-0.65	6.44E-06	-0.75	9.82E-08
PDE7B	1	1	0	0.0	-0.04	8.71E-01	-0.56	6.13E-06
MCC	1	0	1	100.0	-0.41	1.65E-04	-0.53	3.94E-07
ARNT2	1	1	0	0.0	-0.53	8.12E-08	-0.52	1.29E-07
TMEM119	1	0	1	100.0	-0.73	5.86E-06	-0.74	3.75E-06
IFIT1	1	1	0	0.0	0.00	9.96E-01	-0.56	4.08E-04
IRS2	1	1	0	0.0	-0.56	4.26E-05	-0.58	1.65E-05
FNIP1	1	0	1	100.0	-0.66	5.95E-08	-0.63	1.65E-07

References

1 Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J. et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921

2 Klein, S. J., O'Neill, R. J. (2018) Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Res* **26**, 5-23

3 Feschotte, C., Pritham, E. J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* **41**, 331-68

4 Ponicsan, S. L., Kugel, J. F., Goodrich, J. A. (2010) Genomic gems: SINE RNAs regulate mRNA production. *Curr Opin Genet Dev* **20**, 149-55

5 Malik, H. S., Henikoff, S., Eickbush, T. H. (2000) Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res* **10**, 1307-18

6 Boeke, J. D. a. S., J. P. (1997) Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements. *In Retroviruses; Coffin, J.M., Hughes, S.H. and Varmus, H.E. Eds* **Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press**,

7 Mills, R. E., Bennett, E. A., Iskow, R. C., Devine, S. E. (2007) Which transposable elements are active in the human genome? *Trends Genet* **23**, 183-91

8 Cordaux, R., Batzer, M. A. (2009) The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**, 691-703

9 Perez-Stable, C., Ayres, T. M., Shen, C. K. (1984) Distinctive sequence organization and functional programming of an Alu repeat promoter. *Proc Natl Acad Sci U S A* **81**, 5291-5

10 Fuhrman, S. A., Deininger, P. L., LaPorte, P., Friedmann, T., Geiduschek, E. P. (1981) Analysis of transcription of the human Alu family ubiquitous repeating element by eukaryotic RNA polymerase III. *Nucleic Acids Res* **9**, 6439-56

11 Cantarella, S., Carnevali, D., Morselli, M., Conti, A., Pellegrini, M., Montanini, B. et al. (2019) Alu RNA Modulates the Expression of Cell Cycle Genes in Human Fibroblasts. *Int J Mol Sci* **20**,

12 Wei, W., Gilbert, N., Ooi, S. L., Lawler, J. F., Ostertag, E. M., Kazazian, H. H. et al. (2001) Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol* **21**, 1429-39

13 Khazina, E., Truffault, V., Buttner, R., Schmidt, S., Coles, M., Weichenrieder, O. (2011) Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat Struct Mol Biol* **18**, 1006-14

14 Weichenrieder, O., Repanas, K., Perrakis, A. (2004) Crystal structure of the targeting endonuclease of the human LINE-1 retrotransposon. *Structure* **12**, 975-86

15 Mathias, S. L., Scott, A. F., Kazazian, H. H., Jr., Boeke, J. D., Gabriel, A. (1991) Reverse transcriptase encoded by a human transposable element. *Science* **254**, 1808-10

16 Alisch, R. S., Garcia-Perez, J. L., Muotri, A. R., Gage, F. H., Moran, J. V. (2006) Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev* **20**, 210-24

17 Basame, S., Wai-lun Li, P., Howard, G., Branciforte, D., Keller, D., Martin, S. L. (2006) Spatial assembly and RNA binding stoichiometry of a LINE-1 protein essential for retrotransposition. *J Mol Biol* **357**, 351-7

18 Dai, L., LaCava, J., Taylor, M. S., Boeke, J. D. (2014) Expression and detection of LINE-1 ORF-encoded proteins. *Mob Genet Elements* **4**, e29319

19 Cost, G. J., Feng, Q., Jacquier, A., Boeke, J. D. (2002) Human L1 element target-primed reverse transcription in vitro. *EMBO J* **21**, 5899-910

20 Singer, T., McConnell, M. J., Marchetto, M. C., Coufal, N. G., Gage, F. H. (2010) LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends Neurosci* **33**, 345-54

21 Feng, Q., Moran, J. V., Kazazian, H. H., Jr., Boeke, J. D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905-16

22 Naufer, M. N., Furano, A. V., Williams, M. C. (2019) Protein-nucleic acid interactions of LINE-1 ORF1p. *Semin Cell Dev Biol* **86**, 140-149

23 Cordaux, R., Hedges, D. J., Herke, S. W., Batzer, M. A. (2006) Estimating the retrotransposition rate of human Alu elements. *Gene* **373**, 134-7

24 Bennett, E. A., Keller, H., Mills, R. E., Schmidt, S., Moran, J. V., Weichenrieder, O. et al. (2008) Active Alu retrotransposons in the human genome. *Genome Res* **18**, 1875-83

25 Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V. et al. (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* **100**, 5280-5

26 Dewannieux, M., Esnault, C., Heidmann, T. (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* **35**, 41-8

27 Smith, B. L., Gallie, D. R., Le, H., Hansma, P. K. (1997) Visualization of poly(A)-binding protein complex formation with poly(A) RNA using atomic force microscopy. *J Struct Biol* **119**, 109-17

28 Dewannieux, M., Heidmann, T. (2005) Role of poly(A) tail length in Alu retrotransposition. *Genomics* **86**, 378-81

29 Roy-Engel, A. M., Salem, A. H., Oyeniran, O. O., Deininger, L., Hedges, D. J., Kilroy, G. E. et al. (2002) Active Alu element "A-tails": size does matter. *Genome Res* **12**, 1333-44

30 Carnevali, D., Dieci G. (2015) Alu expression profiles as a novel RNA signature in biology and disease. *RNA & DISEASE* **2**, 1-4

31 Conti, A., Carnevali, D., Bollati, V., Fustinoni, S., Pellegrini, M., Dieci, G. (2015) Identification of RNA polymerase III-transcribed Alu loci by computational screening of RNA-Seq data. *Nucleic Acids Res* **43**, 817-35

32 Comeaux, M. S., Roy-Engel, A. M., Hedges, D. J., Deininger, P. L. (2009) Diverse cis factors controlling Alu retrotransposition: what causes Alu elements to die? *Genome Res* **19**, 545-55

33 Liu, W. M., Schmid, C. W. (1993) Proposed roles for DNA methylation in Alu transcriptional repression and mutational inactivation. *Nucleic Acids Res* **21**, 1351-9

34 Chesnokov, I., Schmid, C. W. (1996) Flanking sequences of an Alu source stimulate transcription in vitro by interacting with sequence-specific transcription factors. *J Mol Evol* **42**, 30-6

35 Li, T. H., Schmid, C. W. (2001) Differential stress induction of individual Alu loci: implications for transcription and retrotransposition. *Gene* **276**, 135-41

36 Roy, A. M., West, N. C., Rao, A., Adhikari, P., Aleman, C., Barnes, A. P. et al. (2000) Upstream flanking sequences and transcription of SINEs. *J Mol Biol* **302**, 17-25

37 Aleman, C., Roy-Engel, A. M., Shaikh, T. H., Deininger, P. L. (2000) Cis-acting influences on Alu RNA levels. *Nucleic Acids Res* **28**, 4755-61

38 Calabrese, P. P., Durrett, R. T., Aquadro, C. F. (2001) Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation models. *Genetics* **159**, 839-52

39 Belancio, V. P., Roy-Engel, A. M., Pochampally, R. R., Deininger, P. (2010) Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res* **38**, 3909-22

40 Deininger, P. (2011) Alu elements: know the SINEs. *Genome Biol* **12**, 236

41 Kroutter, E. N., Belancio, V. P., Wagstaff, B. J., Roy-Engel, A. M. (2009) The RNA polymerase dictates ORF1 requirement and timing of LINE and SINE retrotransposition. *PLoS Genet* **5**, e1000458

42 Smit, A. F. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev* **9**, 657-63

43 Quentin, Y. (1994) Emergence of master sequences in families of retroposons derived from 7sl RNA. *Genetica* **93**, 203-15

44 Ullu, E., Tschudi, C. (1984) Alu sequences are processed 7SL RNA genes. *Nature* **312**, 171-2

45 Kriegs, J. O., Churakov, G., Jurka, J., Brosius, J., Schmitz, J. (2007) Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends Genet* **23**, 158-61

46 Quentin, Y. (1992) Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. *Nucleic Acids Res* **20**, 487-93

47 Konkel, M. K., Walker, J. A., Batzer, M. A. (2010) LINEs and SINEs of primate evolution. *Evol Anthropol* **19**, 236-249

48 Schramm, L., Hernandez, N. (2002) Recruitment of RNA polymerase III to its target promoters. *Genes Dev* **16**, 2593-620

49 White, R. J. (2011) Transcription by RNA polymerase III: more complex than we thought. *Nat Rev Genet* **12**, 459-63

50 Zhang, X. O., Gingeras, T. R., Weng, Z. (2019) Genome-wide analysis of polymerase III-transcribed Alu elements suggests cell-type-specific enhancer function. *Genome Res*,

51 Oler, A. J., Traina-Dorge, S., Derbes, R. S., Canella, D., Cairns, B. R., Roy-Engel, A. M. (2012) Alu expression in human cell lines and their retrotranspositional potential. *Mob DNA* **3**, 11

52 van Arensbergen, J., FitzPatrick, V. D., de Haas, M., Pagie, L., Sluimer, J., Bussemaker, H. J. et al. (2017) Genome-wide mapping of autonomous promoter activity in human cells. *Nat Biotechnol* **35**, 145-153

53 Shaikh, T. H., Roy, A. M., Kim, J., Batzer, M. A., Deininger, P. L. (1997) cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. *J Mol Biol* **271**, 222-34

54 Paulson, K. E., Schmid, C. W. (1986) Transcriptional inactivity of Alu repeats in HeLa cells. *Nucleic Acids Res* **14**, 6145-58

55 Oler, A. J., Alla, R. K., Roberts, D. N., Wong, A., Hollenhorst, P. C., Chandler, K. J. et al. (2010) Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nat Struct Mol Biol* **17**, 620-8

56 Gjidoda, A., Henry, R. W. (2013) RNA polymerase III repression by the retinoblastoma tumor suppressor protein. *Biochim Biophys Acta* **1829**, 385-92

57 Kenneth, N. S., Ramsbottom, B. A., Gomez-Roman, N., Marshall, L., Cole, P. A., White, R. J. (2007) TRRAP and GCN5 are used by c-Myc to activate RNA polymerase III transcription. *Proc Natl Acad Sci U S A* **104**, 14917-22

58 Agarwal, P., Enroth, S., Teichmann, M., Jernberg Wiklund, H., Smit, A., Westermark, B. et al. (2016) Growth signals employ CGGBP1 to suppress transcription of Alu-SINEs. *Cell Cycle* **15**, 1558-71

59 Xie, H., Wang, M., Bonaldo Mde, F., Smith, C., Rajaram, V., Goldman, S. et al. (2009) High-throughput sequence-based epigenomic analysis of Alu repeats in human cerebellum. *Nucleic Acids Res* **37**, 4331-40

60 Choy, M. K., Movassagh, M., Goh, H. G., Bennett, M. R., Down, T. A., Foo, R. S. (2010) Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. *BMC Genomics* **11**, 519

61 Liu, W. M., Maraia, R. J., Rubin, C. M., Schmid, C. W. (1994) Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. *Nucleic Acids Res* **22**, 1087-95

62 Jorda, M., Diez-Villanueva, A., Mallona, I., Martin, B., Lois, S., Barrera, V. et al. (2017) The epigenetic landscape of Alu repeats delineates the structural and functional genomic architecture of colon cancer cells. *Genome Res* **27**, 118-132

63 Daskalos, A., Nikolaidis, G., Xinarianos, G., Savvari, P., Cassidy, A., Zakopoulou, R. et al. (2009) Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int J Cancer* **124**, 81-7

64 Varshney, D., Vavrova-Anderson, J., Oler, A. J., Cowling, V. H., Cairns, B. R., White, R. J. (2015) SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. *Nat Commun* **6**, 6569

65 Li, W., Notani, D., Rosenfeld, M. G. (2016) Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* **17**, 207-23

66 Shlyueva, D., Stampfel, G., Stark, A. (2014) Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* **15**, 272-86

67 Su, M., Han, D., Boyd-Kirkup, J., Yu, X., Han, J. J. (2014) Evolution of Alu elements toward enhancers. *Cell Rep* **7**, 376-385

68 Ward, M. C., Wilson, M. D., Barbosa-Morais, N. L., Schmidt, D., Stark, R., Pan, Q. et al. (2013) Latent regulatory potential of human-specific repetitive elements. *Mol Cell* **49**, 262-72

69 Kondo, Y., Issa, J. P. (2003) Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. *J Biol Chem* **278**, 27658-62

70 De Cecco, M., Criscione, S. W., Peckham, E. J., Hillenmeyer, S., Hamm, E. A., Manivannan, J. et al. (2013) Genomes of replicatively senescent cells undergo global

epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell* **12**, 247-56

71 Li, T. H., Kim, C., Rubin, C. M., Schmid, C. W. (2000) K562 cells implicate increased chromatin accessibility in Alu transcriptional activation. *Nucleic Acids Res* **28**, 3031-9

72 Kim, C., Rubin, C. M., Schmid, C. W. (2001) Genome-wide chromatin remodeling modulates the Alu heat shock response. *Gene* **276**, 127-33

73 Liu, W. M., Chu, W. M., Choudary, P. V., Schmid, C. W. (1995) Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res* **23**, 1758-65

74 Singh, K., Carey, M., Saragosti, S., Botchan, M. (1985) Expression of enhanced levels of small RNA polymerase III transcripts encoded by the B2 repeats in simian virus 40-transformed mouse cells. *Nature* **314**, 553-6

75 Larminie, C. G., Sutcliffe, J. E., Tosh, K., Winter, A. G., Felton-Edkins, Z. A., White, R. J. (1999) Activation of RNA polymerase III transcription in cells transformed by simian virus 40. *Mol Cell Biol* **19**, 4927-34

76 Karijolich, J., Abernathy, E., Glaunsinger, B. A. (2015) Infection-Induced Retrotransposon-Derived Noncoding RNAs Enhance Herpesviral Gene Expression via the NF-kappaB Pathway. *PLoS Pathog* **11**, e1005260

77 Karijolich, J., Zhao, Y., Alla, R., Glaunsinger, B. (2017) Genome-wide mapping of infection-induced SINE RNAs reveals a role in selective mRNA export. *Nucleic Acids Res* **45**, 6194-6208

78 Jang, K. L., Latchman, D. S. (1992) The herpes simplex virus immediate-early protein ICP27 stimulates the transcription of cellular Alu repeated sequences by increasing the activity of transcription factor TFIIIC. *Biochem J* **284 (Pt 3)**, 667-73

79 Russanova, V. R., Driscoll, C. T., Howard, B. H. (1995) Adenovirus type 2 preferentially stimulates polymerase III transcription of Alu elements by relieving repression: a potential role for chromatin. *Mol Cell Biol* **15**, 4282-90

80 Chang, D. Y., Hsu, K., Maraia, R. J. (1996) Monomeric scAlu and nascent dimeric Alu RNAs induced by adenovirus are assembled into SRP9/14-containing RNPs in HeLa cells. *Nucleic Acids Res* **24**, 4165-70

81 Panning, B., Smiley, J. R. (1993) Activation of RNA polymerase III transcription of human Alu repetitive elements by adenovirus type 5: requirement for the E1b 58-kilodalton protein and the products of E4 open reading frames 3 and 6. *Mol Cell Biol* **13**, 3231-44

82 Panning, B., Smiley, J. R. (1995) Activation of expression of multiple subfamilies of human Alu elements by adenovirus type 5 and herpes simplex virus type 1. *J Mol Biol* **248**, 513-24

83 Datta, S., Soong, C. J., Wang, D. M., Harter, M. L. (1991) A purified adenovirus 289-amino-acid E1A protein activates RNA polymerase III transcription in vitro and alters transcription factor TFIIIC. *J Virol* **65**, 5297-304

84 Hoeffler, W. K., Kovelman, R., Roeder, R. G. (1988) Activation of transcription factor IIIC by the adenovirus E1A protein. *Cell* **53**, 907-20

85 Yoshinaga, S., Dean, N., Han, M., Berk, A. J. (1986) Adenovirus stimulation of transcription by RNA polymerase III: evidence for an E1A-dependent increase in transcription factor IIIC concentration. *EMBO J* **5**, 343-54

86 Ferrari, R., Gou, D., Jawdekar, G., Johnson, S. A., Nava, M., Su, T. et al. (2014) Adenovirus small E1A employs the lysine acetylases p300/CBP and tumor suppressor Rb to repress select host genes and promote productive virus infection. *Cell Host Microbe* **16**, 663-76

87 Ferrari, R., Pellegrini, M., Horwitz, G. A., Xie, W., Berk, A. J., Kurdistani, S. K. (2008) Epigenetic reprogramming by adenovirus e1a. *Science* **321**, 1086-8

88 Ferrari, R., Su, T., Li, B., Bonora, G., Oberai, A., Chan, Y. et al. (2012) Reorganization of the host epigenome by a viral oncogene. *Genome Res* **22**, 1212-21

89 Horwitz, G. A., Zhang, K., McBrian, M. A., Grunstein, M., Kurdistani, S. K., Berk, A. J. (2008) Adenovirus small e1a alters global patterns of histone modification. *Science* **321**, 1084-5

90 Vetrini, F., Ng, P. (2010) Gene therapy with helper-dependent adenoviral vectors: current advances and future perspectives. *Viruses* **2**, 1886-917

91 Berk., A. J. (2013) Chapter 55: Adenoviridae. *In Fields Virology*; Knipe, D. M., Howley P., Eds**.** Wolters Kluwer Health/Lippincott Williams & Wilkins, 6th edition.

92 Montell, C., Courtois, G., Eng, C., Berk, A. (1984) Complete transformation by adenovirus 2 requires both E1A proteins. *Cell* **36**, 951-61

93 Avvakumov, N., Wheeler, R., D'Halluin, J. C., Mymryk, J. S. (2002) Comparative sequence analysis of the largest E1A proteins of human and simian adenoviruses. *J Virol* **76**, 7968-75

94 Shenk, T. (1996) Adenoviridae: the viruses and their replication. *In Fundamental Virology; Fields, B. N., Knipe, D. M., Howley P. M. Eds* **Lippincott-Raven Publishers, Philadelphia, Pa.**, pp 979–1016

95 Berk, A. J. (2005) Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus. *Oncogene* **24**, 7673-85

96 Pelka, P., Ablack, J. N., Fonseca, G. J., Yousef, A. F., Mymryk, J. S. (2008) Intrinsic structural disorder in adenovirus E1A: a viral molecular hub linking multiple diverse processes. *J Virol* **82**, 7252-63

97 Liu, X., Marmorstein, R. (2007) Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor. *Genes Dev* **21**, 2711-6

98 Frisch, S. M., Mymryk, J. S. (2002) Adenovirus-5 E1A: paradox and paradigm. *Nat Rev Mol Cell Biol* **3**, 441-52

99 Wang, H. G., Rikitake, Y., Carter, M. C., Yaciuk, P., Abraham, S. E., Zerler, B. et al. (1993) Identification of specific adenovirus E1A N-terminal residues critical to the binding of cellular proteins and to the control of cell growth. *J Virol* **67**, 476-88

100 Ferreon, J. C., Martinez-Yamout, M. A., Dyson, H. J., Wright, P. E. (2009) Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proc Natl Acad Sci U S A* **106**, 13260-5

101 Goodman, R. H., Smolik, S. (2000) CBP/p300 in cell growth, transformation, and development. *Genes Dev* **14**, 1553-77

102 O'Connor, M. J., Zimmermann, H., Nielsen, S., Bernard, H. U., Kouzarides, T. (1999) Characterization of an E1A-CBP interaction defines a novel transcriptional adapter motif (TRAM) in CBP/p300. *J Virol* **73**, 3574-81

103 Hamamori, Y., Sartorelli, V., Ogryzko, V., Puri, P. L., Wu, H. Y., Wang, J. Y. et al. (1999) Regulation of histone acetyltransferases p300 and PCAF by the bHLH protein twist and adenoviral oncoprotein E1A. *Cell* **96**, 405-13

104 Ferrari, R., Berk, A. J., Kurdistani, S. K. (2009) Viral manipulation of the host epigenome for oncogenic transformation. *Nat Rev Genet* **10**, 290-4

105 Fuchs, M., Gerber, J., Drapkin, R., Sif, S., Ikura, T., Ogryzko, V. et al. (2001) The p400 complex is an essential E1A transformation target. *Cell* **106**, 297-307

106 Jeong, K. W., Kim, K., Situ, A. J., Ulmer, T. S., An, W., Stallcup, M. R. (2011) Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation. *Nat Struct Mol Biol* **18**, 1358-65

107 Altaf, M., Auger, A., Monnet-Saksouk, J., Brodeur, J., Piquet, S., Cramet, M. et al. (2010) NuA4-dependent acetylation of nucleosomal histones H4 and H2A directly stimulates incorporation of H2A.Z by the SWR1 complex. *J Biol Chem* **285**, 15966-77

108 Svotelis, A., Gevry, N., Gaudreau, L. (2009) Regulation of gene expression and cellular proliferation by histone H2A.Z. *Biochem Cell Biol* **87**, 179-88

109 van Attikum, H., Gasser, S. M. (2005) The histone code at DNA breaks: a guide to repair? *Nat Rev Mol Cell Biol* **6**, 757-65

110 Calo, E., Wysocka, J. (2013) Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**, 825-37

111 Martens, J. A., Winston, F. (2003) Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Curr Opin Genet Dev* **13**, 136-42

112 Rajagopalan, D., Tirado-Magallanes, R., Bhatia, S. S., Teo, W. S., Sian, S., Hora, S. et al. (2018) TIP60 represses activation of endogenous retroviral elements. *Nucleic Acids Res* **46**, 9456-9470

113 Helgason, G. V., O'Prey, J., Ryan, K. M. (2010) Oncogene-induced sensitization to chemotherapy-induced death requires induction as well as deregulation of E2F1. *Cancer Res* **70**, 4074-80

114 Vijayalingam, S., Subramanian, T., Zhao, L. J., Chinnadurai, G. (2016) The Cellular Protein Complex Associated with a Transforming Region of E1A Contains c-MYC. *J Virol* **90**, 1070-9

115 Zhao, L. J., Loewenstein, P. M., Green, M. (2016) Ad E1A 243R oncoprotein promotes association of proto-oncogene product MYC with the NuA4/Tip60 complex via the E1A N-terminal repression domain. *Virology* **499**, 178-184

116 Zhao, L. J., Loewenstein, P. M., Green, M. (2017) Adenovirus E1A TRRAP-targeting domain-mediated enhancement of MYC association with the NuA4 complex activates a panel of MYC target genes enriched for gene expression and ribosome biogenesis. *Virology* **512**, 172-179

117 Chakraborty, A. A., Tansey, W. P. (2009) Adenoviral E1A function through Myc. *Cancer Res* **69**, 6-9

118 Tworkowski, K. A., Chakraborty, A. A., Samuelson, A. V., Seger, Y. R., Narita, M., Hannon, G. J. et al. (2008) Adenovirus E1A targets p400 to induce the cellular oncoprotein Myc. *Proc Natl Acad Sci U S A* **105**, 6103-8

119 Mariner, P. D., Walters, R. D., Espinoza, C. A., Drullinger, L. F., Wagner, S. D., Kugel, J. F. et al. (2008) Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* **29**, 499-509

120 Yakovchuk, P., Goodrich, J. A., Kugel, J. F. (2009) B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes. *Proc Natl Acad Sci U S A* **106**, 5569-74

121 Pagano, A., Castelnuovo, M., Tortelli, F., Ferrari, R., Dieci, G., Cancedda, R. (2007) New small nuclear RNA gene-like transcriptional units as sources of regulatory transcripts. *PLoS Genet* **3**, e1

122 Policarpi, C., Crepaldi, L., Brookes, E., Nitarska, J., French, S. M., Coatti, A. et al. (2017) Enhancer SINEs Link Pol III to Pol II Transcription in Neurons. *Cell Rep* **21**, 2879-2894

123 Lu J. Y., L. C., Tong Li, Ting Wang, Yafei Yin, Ge Zhan, Ke Zhang,, Michelle Percharde, L. W., Qi Peng, Pixi Yan, Hui Zhang, Xue Han, Xianju Bi, Wen, Shao, Y. H., Zhongyang Wu, Peizhe Wang, Wenzhi Li, Jing Zhang, Zai Chang,, Yingping Hou, P. L., Miguel Ramalho-Santos, Jie Na, Wei Xie, Yujie Sun, Xiaohua, Shen, M. R. (2019) L1 and B1 repeats blueprint the spatial organization 3 of chromatin. *bioRxiv*

124 Ferrari R., L. I. d. L. C., Chiara Di Vona, François Le Dilly, Enrique, Vidal, A. L., Javier Quilez Oliete, Laura Jochem, Erin Cutts, Giorgio Dieci,, Alessandro Vannini, M. T., Susana de la Luna and Miguel Beato (2019) TFIIIC binding to Alu elements controls gene expression via chromatin looping and histone acetylation. *bioRxiv*,

125 Gong, C., Maquat, L. E. (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **470**, 284-8

126 Gong, C., Tang, Y., Maquat, L. E. (2013) mRNA-mRNA duplexes that autoelicit Staufen1-mediated mRNA decay. *Nat Struct Mol Biol* **20**, 1214-20

127 Hasler, J., Strub, K. (2006) Alu RNP and Alu RNA regulate translation initiation in vitro. *Nucleic Acids Res* **34**, 2374-85

128 Rubin, C. M., Kimura, R. H., Schmid, C. W. (2002) Selective stimulation of translational expression by Alu RNA. *Nucleic Acids Res* **30**, 3253-61

129 Chen, L. L., Yang, L. (2017) ALUternative Regulation for Gene Expression. *Trends Cell Biol* **27**, 480-490

130 Ivanova, E., Berger, A., Scherrer, A., Alkalaeva, E., Strub, K. (2015) Alu RNA regulates the cellular pool of active ribosomes by targeted delivery of SRP9/14 to 40S subunits. *Nucleic Acids Res* **43**, 2874-87

131 Weichenrieder, O., Stehlin, C., Kapp, U., Birse, D. E., Timmins, P. A., Strub, K. et al. (2001) Hierarchical assembly of the Alu domain of the mammalian signal recognition particle. *RNA* **7**, 731-40

132 Weichenrieder, O., Wild, K., Strub, K., Cusack, S. (2000) Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature* **408**, 167-73

133 Chu, W. M., Ballard, R., Carpick, B. W., Williams, B. R., Schmid, C. W. (1998) Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. *Mol Cell Biol* **18**, 58-68

134 Chang, Y. H., Lau, K. S., Kuo, R. L., Horng, J. T. (2017) dsRNA Binding Domain of PKR Is Proteolytically Released by Enterovirus A71 to Facilitate Viral Replication. *Front Cell Infect Microbiol* **7**, 284

135 Wang, W., Wang, W. H., Azadzoi, K. M., Dai, P., Wang, Q., Sun, J. B. et al. (2016) Alu RNA accumulation in hyperglycemia augments oxidative stress and impairs eNOS and SOD2 expression in endothelial cells. *Mol Cell Endocrinol* **426**, 91-100

136 Wu, J., Chen, Z. J. (2014) Innate immune sensing and signaling of cytosolic nucleic acids. *Annu Rev Immunol* **32**, 461-88

137 Dong, X., Feng, H., Sun, Q., Li, H., Wu, T. T., Sun, R. et al. (2010) Murine gamma-herpesvirus 68 hijacks MAVS and IKKbeta to initiate lytic replication. *PLoS Pathog* **6**, e1001001

138 Tarallo, V., Hirano, Y., Gelfand, B. D., Dridi, S., Kerur, N., Kim, Y. et al. (2012) DICER1 loss and Alu RNA induce age-related macular degeneration via the NLRP3 inflammasome and MyD88. *Cell* **149**, 847-59

139 Kaneko, H., Dridi, S., Tarallo, V., Gelfand, B. D., Fowler, B. J., Cho, W. G. et al. (2011) DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature* **471**, 325-30

140 Hung, T., Pratt, G. A., Sundararaman, B., Townsend, M. J., Chaivorapol, C., Bhangale, T. et al. (2015) The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science* **350**, 455-9

141 Di Ruocco, F., Basso, V., Rivoire, M., Mehlen, P., Ambati, J., De Falco, S. et al. (2018) Alu RNA accumulation induces epithelial-to-mesenchymal transition by modulating miR-566 and is associated with cancer progression. *Oncogene* **37**, 627-637

142 Morales-Hernandez, A., Gonzalez-Rico, F. J., Roman, A. C., Rico-Leo, E., Alvarez-Barrientos, A., Sanchez, L. et al. (2016) Alu retrotransposons promote differentiation of human carcinoma cells through the aryl hydrocarbon receptor. *Nucleic Acids Res* **44**, 4665-83

143 Petrie, J. L., Swan, C., Ingram, R. M., Frame, F. M., Collins, A. T., Dumay-Odelot, H. et al. (2019) Effects on prostate cancer cells of targeting RNA polymerase III. *Nucleic Acids Res* **47**, 3937-3956

144 Castelnuovo, M., Massone, S., Tasso, R., Fiorino, G., Gatti, M., Robello, M. et al. (2010) An Alu-like RNA promotes cell differentiation and reduces malignancy of human neuroblastoma cells. *FASEB J* **24**, 4033-46

145 Jones, N., Shenk, T. (1979) An adenovirus type 5 early gene function regulates expression of other early viral genes. *Proc Natl Acad Sci U S A* **76**, 3665-9

146 Laughlin, C. A., Jones, N., Carter, B. J. (1982) Effect of deletions in adenovirus early region 1 genes upon replication of adeno-associated virus. *J Virol* **41**, 868-76

147 Carnevali, D., Dieci, G. (2017) Identification of RNA Polymerase III-Transcribed SINEs at Single-Locus Resolution from RNA Sequencing Data. *Noncoding RNA* **3**,

148 Flinterman, M. B., Mymryk, J. S., Klanrit, P., Yousef, A. F., Lowe, S. W., Caldas, C. et al. (2007) p400 function is required for the adenovirus E1A-mediated suppression of EGFR and tumour cell killing. *Oncogene* **26**, 6863-74

149 Samuelson, A. V., Narita, M., Chan, H. M., Jin, J., de Stanchina, E., McCurrach, M. E. et al. (2005) p400 is required for E1A to promote apoptosis. *J Biol Chem* **280**, 21915-23

150 Boshart, M., Gissmann, L., Ikenberg, H., Kleinheinz, A., Scheurlen, W., zur Hausen, H. (1984) A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer. *EMBO J* **3**, 1151-7

151 Hoppe-Seyler, K., Bossler, F., Braun, J. A., Herrmann, A. L., Hoppe-Seyler, F. (2018) The HPV E6/E7 Oncogenes: Key Factors for Viral Carcinogenesis and Therapeutic Targets. *Trends Microbiol* **26**, 158-168

152 Hardy S., K. M., Harris-Stansil T., Dai Y. and Phipps M. L. (1997) Construction of Adenovirus Vectors through Cre-lox Recombination. *JOURNAL OF VIROLOGY* **71**, 1842–1849

153 Harlow, E., Franza, B. R., Jr., Schley, C. (1985) Monoclonal antibodies specific for adenovirus early region 1A proteins: extensive heterogeneity in early region 1A products. *J Virol* **55**, 533-46

154 Orioli, A. (12 March 2010) Novel insights into human RNA polymerase III transcription: Non canonical termination and biogenesis of potential regulatory RNAs. *Ph.D. Thesis*, University of Parma, Parma

155 Brummelkamp, T. R., Bernards, R., Agami, R. (2002) A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550-3

156 Daskalova, E., Baev, V., Rusinov, V., Minkov, I. (2007) 3'UTR-located ALU elements: donors of potential miRNA target sites and mediators of network miRNA-based regulatory interactions. *Evol Bioinform Online* **2**, 103-20

157 Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., Pandolfi, P. P. (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033-8

158 Salmena, L., Poliseno, L., Tay, Y., Kats, L., Pandolfi, P. P. (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146**, 353-8

159 Ji, Y., Xie, M., Lan, H., Zhang, Y., Long, Y., Weng, H. et al. (2013) PRR11 is a novel gene implicated in cell cycle progression and lung cancer. *Int J Biochem Cell Biol* **45**, 645-56

160 Naldini, L., Blomer, U., Gage, F. H., Trono, D., Verma, I. M. (1996) Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. *Proc Natl Acad Sci U S A* **93**, 11382-8

161 Chen, C., Okayama, H. (1987) High-efficiency transformation of mammalian cells by plasmid DNA. *Mol Cell Biol* **7**, 2745-52

162 Sakoda, T., Kaibuchi, K., Kishi, K., Kishida, S., Doi, K., Hoshino, M. et al. (1992) smg/rap1/Krev-1 p21s inhibit the signal pathway to the c-fos promoter/enhancer from c-Ki-ras p21 but not from c-raf-1 kinase in NIH3T3 cells. *Oncogene* **7**, 1705-11

163 Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S. et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21

164 Anders, S., Pyl, P. T., Huber, W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-9

165 Love, M. I., Huber, W., Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550

166 Cubenas-Potts, C., Corces, V. G. (2015) Architectural proteins, transcription, and the three-dimensional organization of the genome. *FEBS Lett* **589**, 2923-30

167 Gomez-Roman, N., Grandori, C., Eisenman, R. N., White, R. J. (2003) Direct activation of RNA polymerase III transcription by c-Myc. *Nature* **421**, 290-4

168 Barlev, N. A., Liu, L., Chehab, N. H., Mansfield, K., Harris, K. G., Halazonetis, T. D. et al. (2001) Acetylation of p53 activates transcription through recruitment of coactivators/histone acetyltransferases. *Mol Cell* **8**, 1243-54

169 Crighton, D., Woiwode, A., Zhang, C., Mandavia, N., Morton, J. P., Warnock, L. J. et al. (2003) p53 represses RNA polymerase III transcription by targeting TBP and inhibiting promoter occupancy by TFIIIB. *EMBO J* **22**, 2810-20

170 Bar-Joseph, Z., Siegfried, Z., Brandeis, M., Brors, B., Lu, Y., Eils, R. et al. (2008) Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proc Natl Acad Sci U S A* **105**, 955-60

171 Sinnett, D., Richer, C., Deragon, J. M., Labuda, D. (1991) Alu RNA secondary structure consists of two independent 7 SL RNA-like folding units. *J Biol Chem* **266**, 8675-8

172 Sakamoto, K., Fordis, C. M., Corsico, C. D., Howard, T. H., Howard, B. H. (1991) Modulation of HeLa cell growth by transfected 7SL RNA and Alu gene sequences. *J Biol Chem* **266**, 3031-8

173 Hu, Q., Tanasa, B., Trabucchi, M., Li, W., Zhang, J., Ohgi, K. A. et al. (2012) DICER- and AGO3-dependent generation of retinoic acid-induced DR2 Alu RNAs regulates human stem cell proliferation. *Nat Struct Mol Biol* **19**, 1168-75

174 Baryakin, D. N., Semenov, D. V., Savelyeva, A. V., Koval, O. A., Rabinov, I. V., Kuligina, E. V. et al. (2013) Alu- and 7SL RNA Analogues Suppress MCF-7 Cell Viability through Modulating the Transcription of Endoplasmic Reticulum Stress Response Genes. *Acta Naturae* **5**, 83-93