UNIVERSITÀ DI PARMA

UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN

"*Biotecnologie e Bioscienze*"

CICLO XXXVII

Analysis of gene expression profiling of Alu induction by adenovirus e1a and TFIIIC commitment in hESCs

Coordinatore:

Chiar.ma Prof. Elena Maestri

Tutore:

Prof. Roberto Ferrari

Dottorando: Marco Vezzoli

Anni Accademici 2021/2022 – 2023/2024

Abstract

*Alu*s represent the most abundant class of repetitive elements in the human genome. Despite their abundance, the transcription of these elements is rare, since they are generally repressed. However, some *Alu*s are transcribed and have epigenetic characteristics of active enhancers. Adenovirus 5 has been reported to induce upregulation of *Alu*s transcription. In this thesis, we demonstrate by RNA-seq that the adenoviral small e1a, whose ability to reorganize the epigenome of the host cell through interaction with p300/CBP and pRb has been characterized, is sufficient to trigger *Alu*s transcription, in virtue of its interaction with the chromatin remodeler p400. We thus investigated the molecular mechanism underlying this p400 dependency. By employing genome wide ChIP-seq, we uncovered that the recruitment of TFIIIB is dependent on the interaction between e1a and p400. The epigenetic characterization of the expressed *Alu*s confirmed their enhancer nature as these elements uphold enrichment in their upstream region for p300/CBP, p400 and the YAP/TAZ factors. Moreover, e1a reprograms the *Alu*s in a state reminiscent of a poised enhancer, with an increase in H3K4 mono-methylation (H3K4me1) and a depletion of H3K27 acetylation (H3K27ac). As *Alu*s could serve as platform for 3D genome organization, during our analysis of genome-wide occupancy of the RNA polymerase III machinery in cell infected by adenoviruses, we identified a cluster of genomic regions associated with genes involved in embryonic development, that were depleted of TFIIIC upon e1a expression (e1a is a known de-differentiation factor). This finding led us to further investigate this and will be described in the second part of this thesis. TFIIIC associating with *Alu*s linked to embryonic genes agrees with the identification of the largest TFIIIC subunit (TFIIIIC220) in several CRISPR-Cas9 screenings performed on hESCs to unveil genes involved in the maintenance of stemness and differentiation capabilities. Human TFIIIC has been shown to deposit H3K18ac on *Alu*s in serum-starved conditions and more recently was shown to have specific HAT activity against H3K18 (normally acetylated by the p300/CBP HATs). As hESCs are routinely grown in serum-free medium we, therefore, hypothesized that TFIIIC could play a role in the deposition of the H3K18ac mark in hESCs. The findings described in this thesis strengthen this hypothesis. The genome-wide comparison of H3K18ac and H3K27ac profiles with the occupancy of p300

performed in hESCs, NECs (partially differentiated cells) and IMR90 (fully differentiated cells) showed that in undifferentiated cells most acetylated regions are not occupied by p300. Thus, we produced ChIP-seq of TFIIIC and detected the regions that were bound by TFIIIC in the absence of p300 in hESCs. In these regions H3K18ac and H3K27ac were still enriched, supporting the idea that TFIIIC might be involved in the acetylation of at least H3K18. Gene Ontology and Cistrome dbToolkit score analysis showed that the TFIIIC-bound regions are involved in stemness maintenance and neurodifferentiation. Thus, we suggest that TFIIIC could play an important role in hESCs biology.

Index

Chapter 1

Introduction

Transposable elements of the human genome

Since the first publication of the human genome sequence by the 'human genome project', it has been observed that approximately half of the genome contains a class of mobile genetic elements named transposons [Figure 1] [1], which has been described to directly contribute to the human genome evolution. Several mechanisms have been unveiled for these elements, such as the acquisition of new functions like the creation of new regulatory networks for gene regulation [2], [3], the direct modulation of the RNA polymerase II (Pol2) [4] and they participation in the folding of the genome in its 3D structure [5].

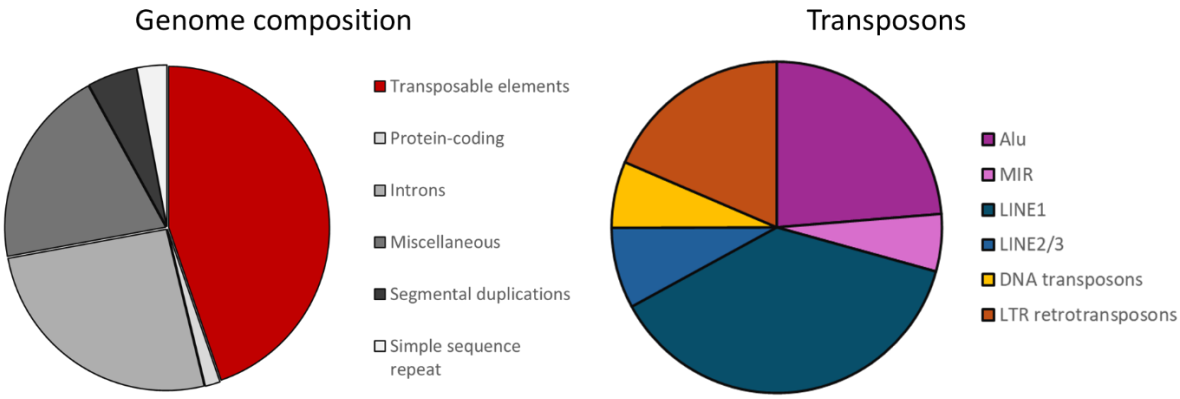

Figure 1. Composition of the human genome. Approximately 45 % of the genome is composed of transposable elements, of which approximately 25 % are *Alu*s [1].

The transposable elements can propagate into the genome through two different mechanisms: the "copy and paste" mechanism that consists of the synthesis of an RNA intermediate that is retrotranscribed and inserted into a different locus (RNA transposon) or the "cut and paste" mechanism in which the element is excised from the locus and inserted into a different one (DNA transposon) [Figure 2] [6].

RNA transposons (or retrotransposons) represent the most abundant transposable elements in the human genome. They are classified into 3 categories: retrovirus-like elements (LTR retrotransposons), Long Interspersed Nuclear Elements (LINEs) and Short Interspersed Nuclear Elements (SINEs) [6].

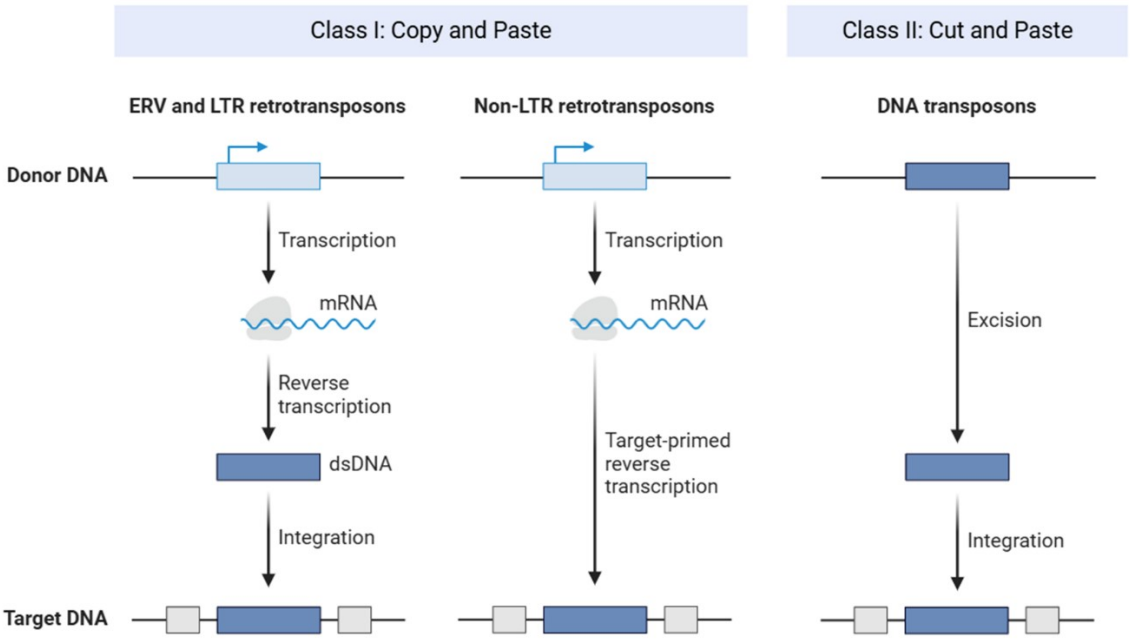

Figure 2. Representation of the two class of transposable element. Class I: the element is transcribed and retrotranscribed. LTR retrotransposons pass through a dsDNA intermediate that is integrated in the genome, whereas non-LTR retrotransposon RNA anneals with the target locus that primes the retrotranscription. Class II: the element is excised and integrated in another locus.

The LTR retrotransposons elements (such as the Human Endogenous Retroviruses HERVs) are derived from ancient retroviruses and characterised by Long Terminal Repeats (LTR) flanking a central region coding for two of the three retrovirus proteins *gag* and *pol* (*env* gene is missing). *Gag* and *pol* allow LTR transposition. LINEs are non-LTR retrotransposon of about 6-8 kbs in length and code for two proteins named ORF1 and ORF2, which are required for the entire retrotransposition process. Therefore, LTRs and LINEs are autonomous retrotransposons. LINEs transcription is carried out by Pol2. On the other hand, SINEs have smaller sizes, about 100-400 bp, transcribed by the Pol3 (Pol3) but lacking coding sequences necessary for their mobility, thus SINEs retrotransposition depends on LINEs-coded ORF1 and ORF2 [Figure 3]. Hence SINEs are non-autonomous retrotransposons [1]. SINEs represent 13 % of the human genome and with ~1'600'000 copies they are the most frequent transposable element in the human genome. Among SINES, *Alu*s represent the largest class of SINEs with ~1'000'000 copies [1].

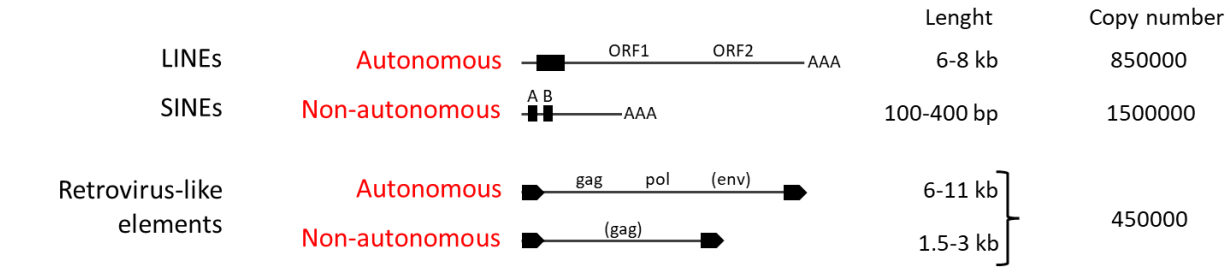

Figure 3. Representation of different classes of retrotransposon. Most transposons fall into one of three classes between LTR, LINE, SINE. Autonomous transposons encode for all components necessary for transposition, unlike non-autonomous ones. Adapted from [1]

Structure and evolution of Alus element

Alus seem to have evolved from the retrotransposition of 7SL RNA (another Pol3-transribed gene), given their homology with its sequence. *Alus* length is about 300 bp and possess a dimeric structure composed of a left arm and a right arm. The left arm harbours a type II Pol3 promoter. A characteristic A_5ATACA_6 sequence separates the two arms with the right arm ending with a poly(A) tail. A canonical or non-canonical Pol3 terminator (4 Ts or a T-rich region respectively), is located upstream of a sequence called 3' trailer unique to each *Alu* RNAs. *Alus* are flanked by Target-site Duplications (TSD) [Figure 4].

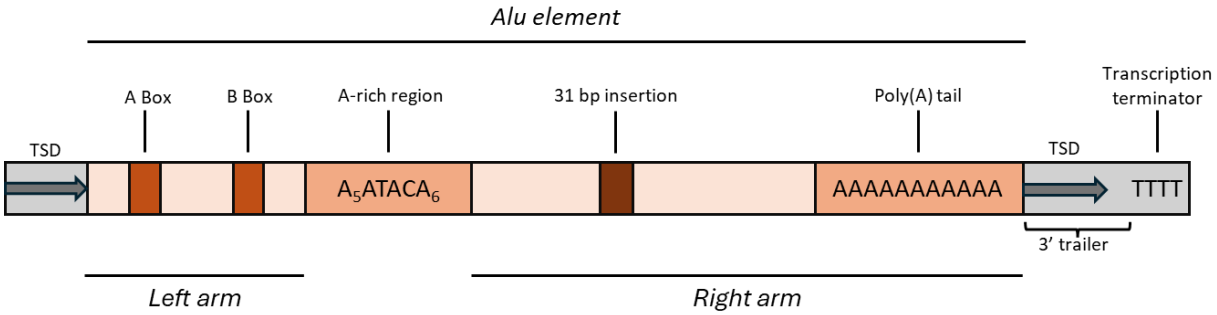

Figure 4. Schematic representation of *Alu*.

The current model of *Alus* evolution is based on the rise of two different lineages from the retrotransposition of a 7SL RNA. One lineage generated Fossil Alu Monomer (FAM) that subsequently evolved into Free Right Alu Monomer (FRAM). The other lineage evolved in the FLAM-C subfamily of the free-left *Alu* monomer [7]. At about 65 million years of age, the two sequences (FRAM and FLAM-C) had a head-to-tail fusion, giving rise to the *Alus* and their subsequent amplification within the genome [8]. *Alus* are classified into three subfamilies:

AluJ, *AluS* and *AluY*. *AluJ* is the most ancient category and contains about 160'000 sequences, all of which are thought to be inactive due to the mutations accumulated. *AluS* is the most common in the human genome with about 551'000 copies and the second oldest family, it also contains some functional *Alu*s. *AluY* is the most recent class and harbours most of the active sequences, with *AluYa5* and *AluYb8* representing the subclasses with the higher transposition activity [9]. To date, the active *Alu*s are estimated to be 852 [9], with a mobilisation rate of 1 for every 20 new births [10].

Transcription of *Alu*s by Pol3

*Alu*s retrotransposition mostly occurs at random genomic loci but often within gene-rich regions [1] : here *Alu*s can be in intragenic or intergenic loci (UTR, exon or intron [11]) an in-sense or antisense compared to the near gene. *Alu*s inserted in intragenic loci with the same orientation of the host gene can therefore be transcribed by Pol2 as a part of the gene transcript, whereas intragenic or antisense *Alu*s are thought to be transcribed exclusively by Pol3. Since the Pol3-derived *Alu* transcripts are involved in the retrotransposition phenomena, while the Pol2-derived *Alu*s lack retrotransposition potential [12], we will focus on that.

The left monomer harbours an A and B box, composing a type II Pol3 promoter. A and B boxes are recognised by the multi-subunit complex Transcription Factor IIIC (TFIIIC), which in turn recruits the Transcription Factor IIIB (TFIIIB) complex on a ~50 bp region upstream of the transcription start site. TFIIIB is the main factor involved in the recruitment of the Pol3 and it is composed of three subunits, namely TBP, BRF1 and BDP1 [13] [Figure 5].

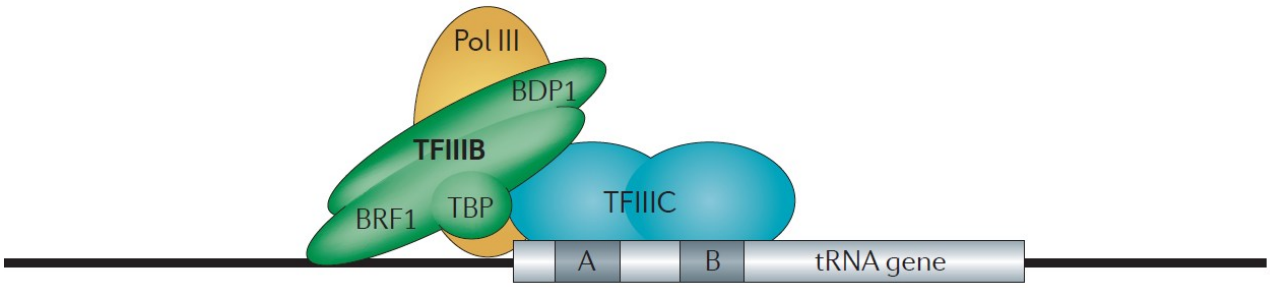

Figure 5. Assembly of Pol3 machinery on Type II promoter. TFIIIC recognize the type II promoter, to recruit TFIIIB, which in turn recruits Pol3. Taken from [14] .

Many transcription factors are shared between Pol2 and Pol3 as the Rb transcription factor [15], or c-myc [16]. The transcription factor AP-1 was found to bind upstream the TSS of *Alu*s and enhances the transcription [11], [17], as well as the transcription factor CEBPB and CTCF [11]. In contrast, CGGBP1 acts as a repressor of transcription in response to growth factors [18]. Interesting CGGBP1 binding site overlaps with the A box, preventing the recruitment of TFIIIC, which is instead recruited during serum starvation [5].

Pol3 transcription in general is tightly regulated in a tissue and cell type-specific manner and differently modulated, depending on cells growth conditions. Therefore, *Alus* transcription will also be subjected to the same strict regulation [11], [19], [20].

Within the cells, abundance of *Alu*s transcripts is generally extremely low due to their weak Pol3 promoter and epigenetic modifications which repress *Alu*s transcription [21]. Studies on HeLa cells have shown that only ~100 loci are active [20] [22], and of these, only 13 depended on Pol3 transcription [23]. Studies on other cell lines or tissue samples emphasised how low *Alu*s transcripts derived from Pol3 is a common theme [11], [24].

*Alu*s epigenetic environment: DNA methylation

DNA methylation occurs mainly in the genome at CpG dinucleotides (CpGs) where a methyl group is added to the fifth position of the cytosine residue, about 25 % of which were estimated to be in *Alu*s [25]. DNA methylation is associated with transcription repression predominantly through two general mechanisms: 1) the modification of the cytosine residue prevents the recognition of the binding site by transcription factors, 2) methylated CpGs recruit the methyl-CpG-binding proteins (MBPs), which interact with transcriptional co-repressor complexes able to remodel the chromatin [26] [Fig. 6].

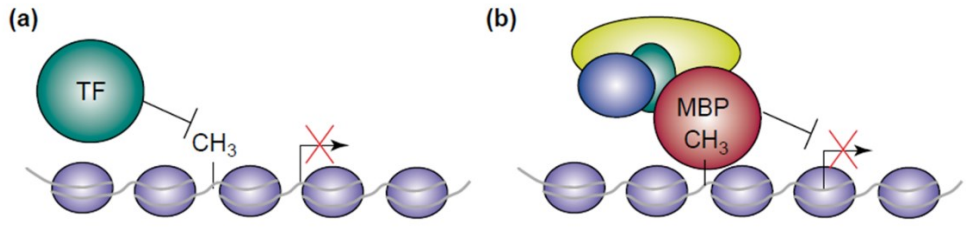

Figure 6. DNA methylation represses transcription. Two main mechanisms of repression mediated by DNA methylation. a) methylation blocks the recruitment of transcription factor. b) methylated DNA recruits MBPs, that recruits co-repressors. Adapted from [26]

*Alu*s are highly methylated. The subclass *AluY* harbours the highest levels of DNA methylation compared to the *AluS* and *AluJ* classes [27]. The involvement of DNA methylation in the repression of *Alu*s transcription is supported by several studies carried out on different systems [28], [29], [30], [31]. However, one study showed that increased *Alu*s transcription was not dependent on DNA demethylation rather by histone modifications [32].

A new hemi-methylation pattern, that correlates with nucleosome positioning, was recently observed on *Alu*s [33]. Here the antisense strand was methylated at two CHG sites (H=A/T/G). Remarkably, the hemi-methylation at CAG sites enhances *Alu*s transcription [33].

*Alu*s epigenetic environment: histone modifications

In addition to DNA methylation, the profile of histone modifications can influence the *Alu*s transcription. A key role in the repression of *Alu*s transcription has been observed for the trimethylation of histone H3-Lysine 9 (H3K9me3). H3K9me3 is associated with heterochromatin where, through the recruitment of HP1, promote the chromatin condensation preventing the binding of transcription factors [34]. The inhibition of SUV39H1 (the enzyme that deposits H3K9me3) in HeLa cells caused a decrease in H3K9me3 at some *Alu*s loci, triggering the recruitment of the Pol3 machinery and hence increasing *Alu*s expression [32].

Transcriptional active *Alu*s are enriched, in a tissue- and cell line-specific manner, with those histone modifications typically associated with active chromatin, such as H3-Lysine 4 mono- and di-methylation (H3K4me1 and H3K4me2), H3-lysine 27 acetylation (H3K27ac), H3-lysine 9 acetylation (H3K9ac), H3-lysine 27 mono-methylation (H3K27me1), H3-lysine 36 tri-methylation (H3K36me3) [19], [35]. Some of these histone modifications are typically found at cis-regulatory regions such as enhancers. H3K4me1 and H3K4me2 contribute to the recruitment of cohesin complex promoting interaction between the enhancer and the target promoter [36]. Moreover, H3K4me1 recruits the SWI/SNF complex [37]. H3K36me3, through the recruitment of Dnmt3a and Dnmt3b, promotes the eRNA transcription [38]. H3K27ac distinguishes active from inactive enhancers [39]. The deposition of H3K27ac at enhancer regions allows the stabilization of BRD4 binding, which promotes the recruitment of the Mediator complex [40]. Despite H3K27ac being a well-recognized marker for active enhancers,

it has been reported that its depletion did not affect the enhancer activity in mouse embryonic stem cells [41].

Therefore, since enhancers are known to be transcribed [42], it is conceivable that these modifications might contribute to the regulation of *Alus*. The transcription factor AP-1 has been observed bound to Pol3-dependent genes [43] and its binding motif to be enriched near the TSS of *Alu*s [19]. AP-1 can recruit p300 (one of the major mammalian histones acetyltransferases [44]), increasing acetylation and promoting transcription [19].

Viral infection activates Alus transcription

*Alu*s transcription, and more generally SINE transcription, is activated in response to many stresses, such as viral infection.

Several studies have reported an increase in SINEs transcription upon infection of mouse cell lines [45], [46], [47] . SV-40 caused increase in B2 SINE (the equivalent of *Alu*s in mice) [45], in the context of a global activation of all RNA polymerases due to the large T antigen of SV-40 [46]. Murine Gammaherpesvirus 68 (MHV68) induced mouse B1 and B2 SINE *in vitro* and *in vivo* to sustain its proliferation [47]. Differently from what was observed for SV40, MHV68 infection did not increase the pol 3 activity but triggered the B2 SINE transcription through serine/threonine protein kinase ORF36. This factor can phosphorylate pRb and, through kinase-independent functions, inhibit HDACs [48].

In line with what has been observed in mice, infections can cause an increase in the expression of SINEs in different human cell lines. Herpex Simplex Virus-1 (HSV-1) induced *Alu*s expression in HeLa and HEK293 cells [49], [50]. As observed for the SV-40 infection, HSV-1 induced Pol3 transcription [51]. The increase in *Alu*s expression in HeLa, HEK293 and fibroblast cell lines has also been observed following the infection with Adenovirus-2 (Ad2) and Adenovirus-5 (Ad5) [52], [53], [54], [55], [56] . This adenoviral induction in human *Alu*s transcription takes place due to the action of the adenoviral protein E1A which sets off a global chromatin remodelling which sustains increased Pol3 transcription [57], [58].

Adenovirus genome and the E1A protein

Adenovirus 5 (Ad5) belongs to the class of dsDNA viruses that infect epithelial cells of the respiratory tract. Ad5 genome is a linear dsDNA with a length of 36 kb divided into five Early Transcription Units (E1A, E1B, E2, E3, E4) that are transcribed before the replication of the genome, four Intermediate transcription units (pIX, Iva2, L4 intermediate and E2 late) and one Late transcription unit under the control of the Major Late Promoter (MLP) that generates 5 different mRNAs (L1, L2, L3, L4, L5)[59] [Figure 7].

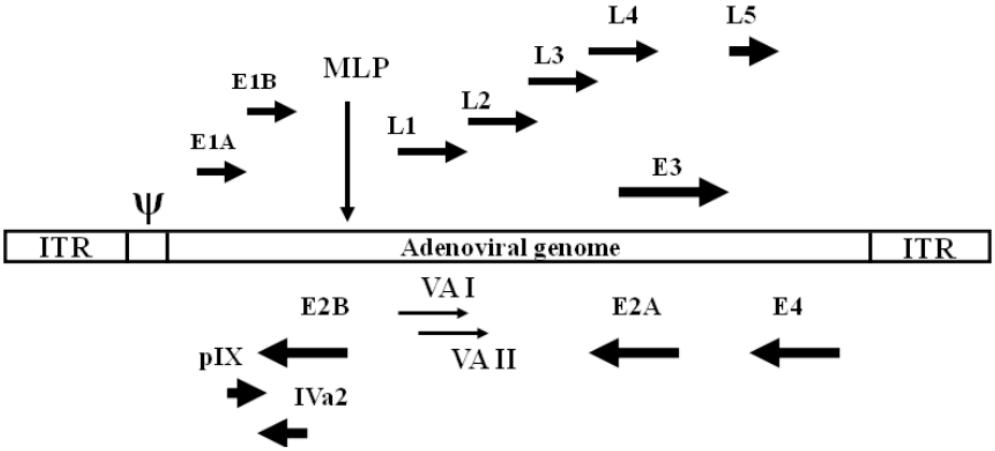

Figure 7. Schematic representation of adenoviral genome.

Once the virus is internalised inside the cell, the viral DNA is delivered into the nucleus, following decapsidation, and the transcription of early genes can start. The first gene transcribed is E1A which activates the transcription of the other genes and interacts with several host proteins to sustain the infection. The proteins coded by the early genes set up the environment for viral replication, inhibiting the immune response (E3), the p53-dependent apoptosis (E1B) and enhancing the protein synthesis by acting on the mTOR signalling pathway (E4) [60].

The E1A pre-mRNA has two different splice sites at the 5', allowing the generation of two mRNAs encoding for a 243 amino acids protein (called: "small e1a" or "e1a") and a 283 amino acids protein (called: "large E1A" or "E1A") [61] [Figure 8].

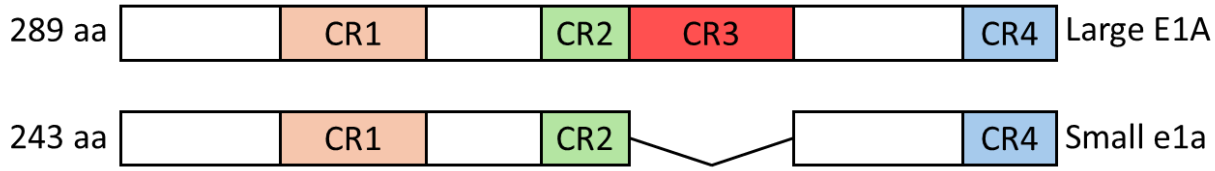

Figure 8. Schematic representation of two splicing variant of E1A. Alternatively splicing of E1A transcript generate two variants. Large E1A is 289 amino acids protein, whereas small e1a is 243 amino acids protein. The latter lack of CR3 domain, responsible for the interaction with MED23.

Large E1A proteins, synthesized by different adenoviral strains that infect primates, have four highly conserved regions, called CR1-CR2-CR3-CR4, that span from residues 42 to 72, 115 to 140, 144 to 191, 251 to 288, respectively [62]. These regions play a pivotal role in the establishment of interaction between E1A and the host proteins pRBs (Rb, p107 and p130), p300/CBP, p400 and the Mediator which are involved in remodelling of the chromatin and regulation of genes key in the induction of S-phase.

Large E1A ability to activate the transcription of viral genes relies on the interaction between the CR3 domain and the MED23 subunit of the Mediator. Lack of CR3 domain in small e1a abolishes that interaction, resulting in the inability to induce the transcription of the viral genes [63], whereas can still induce the S-phase of the cell cycle. Even though E1A is necessary for successful viral DNA replication, it is dispensable for the induction of cell cycle of G1-arrested cells [64].

Interaction e1a-Rb-p300

As mentioned before, e1a induces the S-phase in G_1-arrested cells. This depends on the interactions of e1a with the tumour suppressor Rb and the histone acetyltransferase (HAT) p300 (or EP300). E1a binds Rb through 10 amino acids in the CR1 domain and the L-X-C-X-E motif in the CR2 region [62]. The molecular mechanism that drives oncogenesis depending on the interaction e1a-pRb has been characterized: e1a competes with Rb for the same regions within the transcription factor E2F, since e1a CR2 has higher affinity for E2F than Rb [65], e1a can sequester Rb and then use the CR1 regions to insert an alpha-helix within the same Rb pocket occupied by E2F thereby dissociating the two factors [65]. This clever mechanism abrogates the formation of the E2F-Rb complex and therefore induces host cell cycle and the

transcription of viral genes. In dividing cells preventing the formation of the complex Rb-E2F is dependent on cyclin-CDK complexes-mediated phosphorylation of Rb in response to extracellular mitogenens [66] [Figure 9].

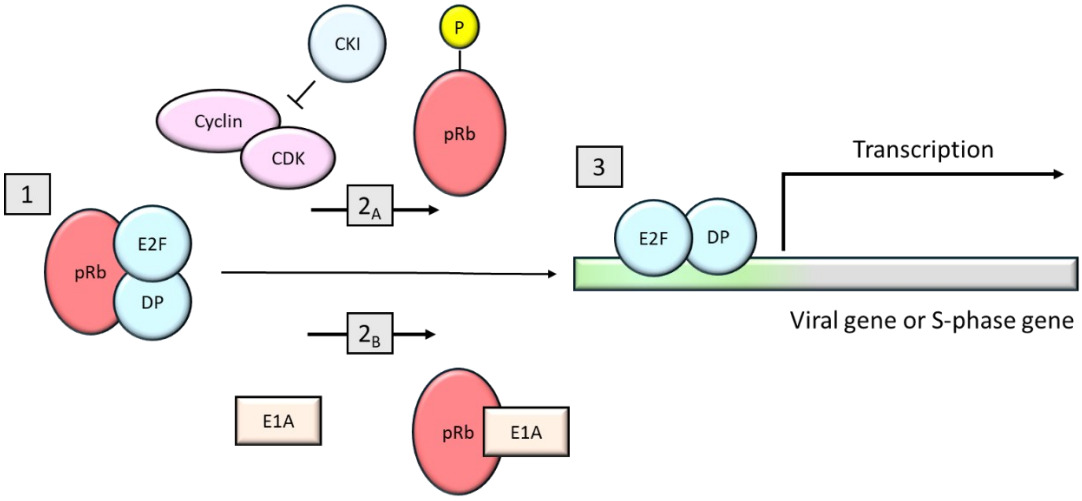

Figure 9. E1A bypass E2F regulation. 1) In G_0-G_1 cells hypophosphorylated Rb sequester E2F, preventing the transcription activation. 2a) Rb is phosphorylated by Cyclin-CDK complexes, causing the release of E2F. 2b) E1A sequester Rb, inhibits its ability to interact with E2F. 3) E2F does not interact with Rb anymore and can activate the transcription of target genes. Adapted from [66].

The release of Rb from E2F, in addition to the enhancement of transcription of genes involved in cell proliferation, causes an increment in the availability of this protein, which is accumulated at e1a-repressed genes involved in the immune response, differentiation and synthesis of the extracellular matrix [67], [68] [Figure 12a].

The residues 1-25 and 54-83 (localized inside the CR1) allow the interaction of e1a with the HAT p300 [69]. p300 is a transcriptional coactivator that, through the associations with other transcription factors, can be recruited at gene promoters [70] and acetylates lysine in the histone tails or on other transcription factors, using its intrinsic acetyltransferase activity. The interaction between e1a and the TRAM (transcriptional adaptor motif) domain of p300 inhibits the HAT activity, leading to a restricted residual acetyl-transferase activity to specific genomic loci [Figure 10].

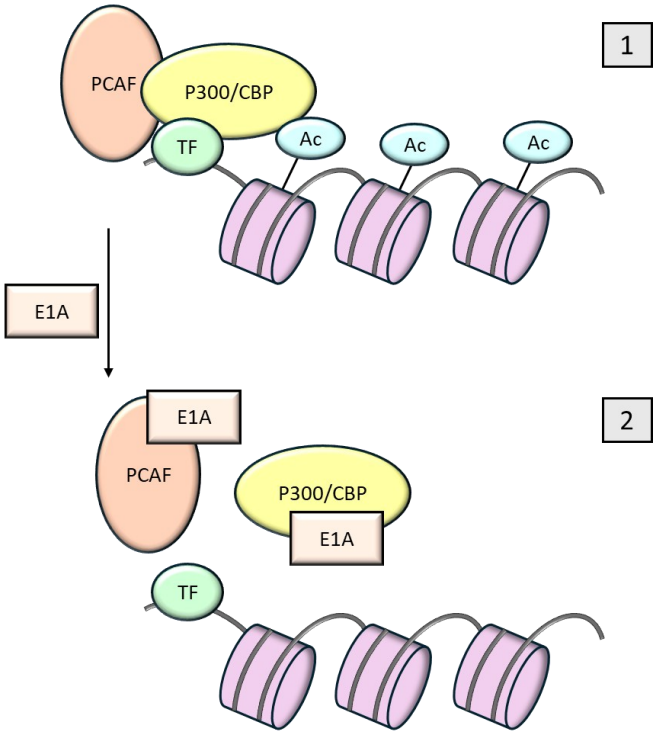

Figure 10. E1A inhibits p300 HAT activity. 1) p300 is recruited by transcription factors and acetylates H3K18 and H3K27. 2) E1A prevents the recruitment of p300 and inhibits its HAT activity. Adapted from[66] .

Going into more detail, the interaction e1a-p300 causes a global H3K27 and H3K18 deacetylation, however H3K18ac is found increased at cell cycle genes [67], [71]. Small e1a therefore leads the transcription of genes involved in cell proliferation removing the endogenous inhibitor (Rb) and causing changes in the epigenetic landscape of the E2F-regulated genes through the preserved HAT activity of p300 [72] [Figure 11].

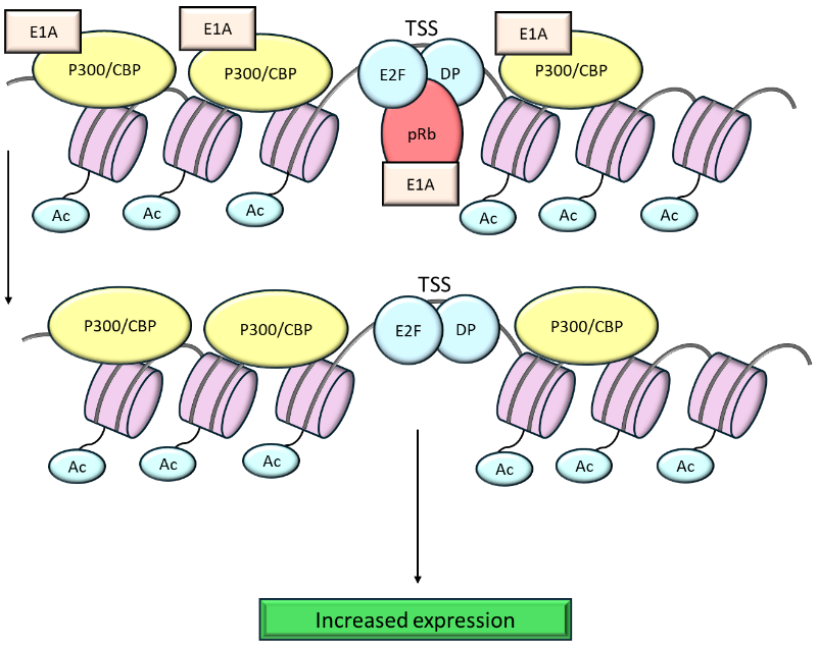

Figure 11. Schematic representation of E1A induced transcription. The activation of cell-cycle genes transcription depends on the simultaneously remotion of pRb from E2F and deposition of H3K18ac by p300 caused by E1A.

E1a represses the transcription of those genes involved in differentiation and immune response. Two different mechanisms, involving p300, Rb and two Rb-like proteins (p107 and p130), have been identified. The first mechanism involves the recruitment of p107 and e1a on genes promoting cellular differentiation [66] [Figure 12a], whereas the second mechanism depends on the recruitment of a complex formed by p300-e1a-Rb and p130 on the antiviral response genes [65], [66]. The hypoacetylation caused by the inhibition of the HAT activity of p300 by e1a, in conjunction with the acetylation of Rb, participates in chromatin condensation at these loci, repressing their transcription [67] [Figure 12b].

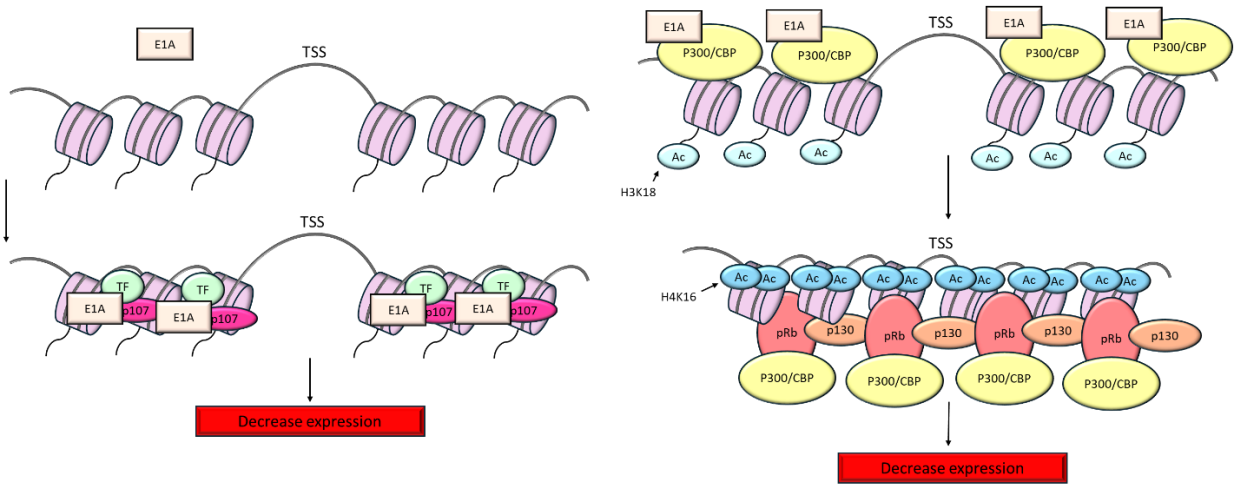

Figure 12. Schematic representation of two models of E1A induced repression. a) Rb (p107) is redistributed to repressed genes after the release from E2F. **b)** The formation of a ternary complex formed by p300-e1a-Rb causes hypoacetylation and chromatin condensation.

In conclusion, the interactions that are established between e1a, Rb and p300/CBP led to a global change in the epigenetic landscape of the host cell to sustain an efficient viral replication.

Interaction e1a-p400

As mentioned before, e1a binds p400, homologous to the yeast SWI2/SNF2 nucleosome remodelling complex, through the residues 26-35 and this interaction is necessary to cause cell transformation [73]. Small e1a interacts in two different ways with p400: indirectly through interaction with TRAPP and directly in a competitive manner with TAP54 [66] [Figure 13].

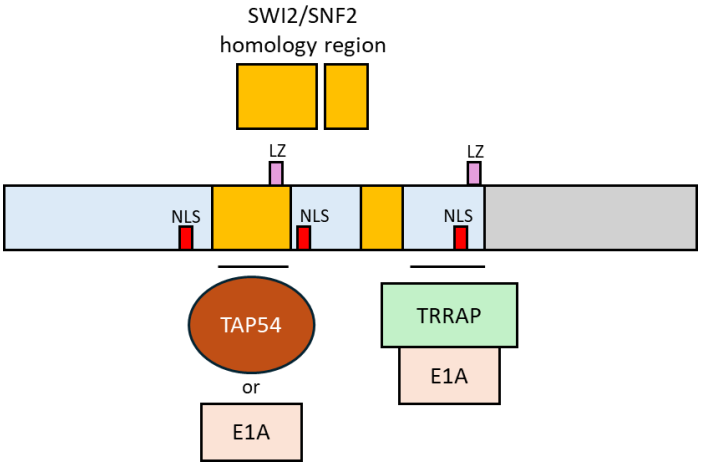

Figure 13. Schematic representation of primary structure of p400. Adapted from [66]

P400 and TRAPP have been identified also in other complexes with HAT activity, such as the TIP60/p400: it consists of 17 subunits with a molecular weight of about 1.7 Megadalton with p400 that in addition to having the remodelling activity, represents the scaffold of the entire complex [74], [75] [Figure 14]. The TIP60 complex is involved in the maintenance of chromatin accessibility where TIP60 can be recruited by H3K4me1, an enhancer mark [76], and it is involved in the acetylation and deposition of H2A.Z, a histone variant that makes the chromatin more accessible [77], [78].

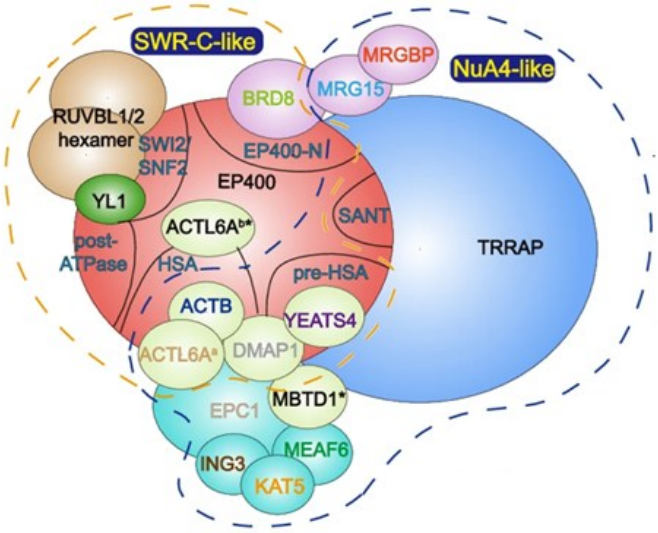

Figure 14. Representation of TIP0 complex. TIP60 complex is the unique complex that exhibit chromatin remodelling and histone-modifying activities. The ability to replace H2A with H2A.Z rely on SWR-C-like module, while the HAT activity resides in the NuA4-like module. Adapted from [74].

The TIP60/p400 complex is also involved in heterochromatinization and repression of transcription [79], [80]. TIP60 has been also observed to be responsible for silencing HERV elements [81].

The mechanism that explains the involvement of the interaction between e1a and cell transformation is not fully understood, and several hypotheses have been proposed. It has been reported that p400 bound and inhibited the transcription factor E2F. Hence, e1a could sequester p400 away from E2F, in a manner like observed with Rb [82]. The interaction e1a-p400, with p300 and Rb, could be part of a multiprotein complex that can lead to chromatin condensation [67]. The complex e1a-p400 can also inhibit the ubiquitination of the

transcription factor Myc, promoting the interaction between p400 and Myc at Myc-target gene promoters and consequently activating their transcription [83].

Less is known about the role of the interaction between e1a and p400 compared to the one with Rb and p300, however, it appears that the formation of this complex, along with other proteins such as TIP60, may also lead to a modification of the cellular epigenetic profile to support viral replication.

Aim of the research

*Alu*s constitute about 11% of the human genome and, although their transcription is generally epigenetically silenced, some elements are actively transcribed in a cell type-dependent manner and possess epigenetic features of active enhancers. *Alu*s transcription is triggered by different cell stresses, such as heat shock [84], [85], cancer progression [86] and virus infection[52], [53], [54], [55], [56]. Adenovirus-5 infection is known to trigger *Alu*s over-expression in different cell lines, but the underlying molecular mechanism is still poorly understood. This phenomenon occurs in a context where the adenoviral protein E1A activates Pol3 transcription possibly acting on TFIIIC. Small e1a causes a cellular epigenetic alteration at promoters and enhancers (which are enriched in *Alu*s [87]), interacting with key chromatin regulators, such as p300/CBP, Rb and p400, and might indeed be a putative candidate for *Alu*s transcription regulation.

We hypothesised that e1a could potentially be regulating Pol3 activity via an epigenome reprogramming of human IMR90 primary fibroblasts, therefore regulating *Alus* expression. To test this, we infected IMR90 cells with Ad-5 replicative mutants, expressing wt small e1a or different small e1a protein uncapable of interacting with their main host partners (p300, Rb and p400). We then performed gene expression profiling (RNA-seq) of cells infected with these mutant (compared to the wt) and captured the *Alu*s expression profile at single locus resolution. We then associated our RNA-seq data with chromatin-immunoprecipitation and sequencing (ChIP-seq) of several Pol3 factors, histone modifications and regulators to fully characterize the regulatory epigenome possibly affecting *Alu*s transcription.

Methods

Cell culture and viruses

IMR90 primary human fetal fibroblasts were purchased from the American Type Culture Collection (ATCC). Cells were grown in Dulbecco's modified medium supplemented with 10% fetal bovine serum (FBS), 100 U/ml penicillin and 100 µg/ml streptomycin at 37°C in 5% CO_2. Ad5 *dl*1500 containing a 9-bp deletion removing the 13S E1A mRNA 5'-splice site was described. The ΔE1A deletion mutant *dl*312 was isolated as previously described. Small e1a binding mutants (e1a_RB_b$^-$, e1a_p300_b$^-$, e1a_p400_b$^-$) were constructed as in, and incorporated in the *dl*1500 background. *dl1500* and e1a binding mutant constructs were cloned into shuttle plasmid pAd-lox. LoxP recombination between HAdV-C5 backbone ψ5 and the shuttle plasmid pAdLox, as well as propagation of viruses, were performed as described in [88].

Virus infection

IMR90 cells were grown to confluence in 60 mm Petri dishes and incubated for two more days in the same medium to arrest cells in G_1/G_0. Before the infection, the medium was collected (conditioned medium) and cells were incubated with mock or the indicated Ad5 mutant for 1 h in PBS. Cells were washed with PBS and transferred back to the conditioned medium for 24 h. To achieve a comparable amount of wild-type and mutant e1a protein expression, the infection was performed at 40 MOI for *dl*1500 and *dl*312, 160 MOI for e1a_RB_b$^-$, 60 MOI for e1a_p300_b$^-$ and 6 MOI for e1a_p400_b$^-$.

Total RNA-seq library preparation

Cells were collected in TriZOL (Thermofisher) and total RNA was extracted with Direct-zol RNA MiniPrep Plus (Zymo research) and treated with DNAseI (Thermofisher); subsequently, the enzyme was inactivated adding 2 mM EDTA and heating at 65°C for 10 minutes.

1 µg of total RNA from cells infected with *dl*1500, *dl*312, or mock-infected was processed using a Ribo-Zero rRNA Removal Kit (Epicentre) to remove the rRNA. The RNA-seq libraries were prepared using the Illumina TruSeq stranded RNA library Preparation kit. The sequencing was performed using a HiSeq4000 Illumina Sequencer, obtaining 100-base-long paired-ends reads, with a sequencing depth of ~60 million reads per sample. The experiment was conducted in duplicate.

1 µg of total RNA from cells infected with *dl*1500, e1a_RB_b^-, e1a_p300_b^-, e1a_p400_b^- or mock-infected was processed using the RiboCop rRNA Depletion Kit (Lexogen). The RNA-seq libraries were prepared using the Illumina TruSeq stranded RNA library Preparation kit. The sequencing was performed using a NovaSeq Illumina sequencer, obtaining 150-base-long paired-ends reads, with a sequencing depth of ~100 million reads per sample. The experiment was conducted in duplicate.

RNA-seq analyses

The alignment of the reads was performed to the GRCh38 human reference genome using STAR [89]. Only uniquely mapped reads were considered for downstream analyses and subjected to counting with the featureCounts tool of the SubRead Python package [90]. The *Alu*s RNA profiling was performed as describe in [91] with some modifications to reduce false positives. These involved the additions of parameters to set the cut-off value for the ratio of the expression coverage between the *Alu* body and its upstream and downstream region and the addition of a parameter to control the fraction of *Alu* body sequence that should be covered by reads to enable the identification of shorter and processed *Alu* transcripts; identified *Alu*s were filtered for an expression coverage of at least 1000 nt, corresponding to 5 pair-ended reads of length 100 nt. The union of these *Alu*s constitutes the whole list of expression-positive *Alu*s (ep*Alu*s). Differentially expressed *Alu*s were identified with DESeq2

package [92]. *Alu*s with $\log_2$ fold-change ≥0.5 or ≤-0.5 and an adjusted p-value < 0.05 were considered differentially expressed.

Alu proximity to protein-coding genes

GENCODE annotation v24 (human genome assembly GRCh38/hg38) was used to analyse *Alu* distance to protein-coding genes. RepeatMasker tracks were downloaded from the UCSC Genome Browser for human genome assembly hg38. *Alus* that are positioned outside of protein-coding genes (intergenic *Alu*s) and *Alu*s that map to introns or exons of annotated genes in an antisense orientation (antisense *Alu*s) were selected to obtain the intergenic/antisense *Alu*s subset (802571 *Alu*s). The expression of intergenic/antisense *Alu*s was considered only for the elements belonging to the single, comprehensive list of ep*Alu*s, whereas the remaining intergenic/antisense annotated *Alu*s were used as a control for unexpressed *Alu*s. The presence of an *Alu* within protein-coding genes was analysed using the "intersectBed" function of the BEDtools program package v2.29.1 [93] and custom R scripts. Enrichment analyses were performed using a two-tailed Fisher's exact test. The transcription start site (TSS) position of protein-coding genes was extracted as the first or last nucleotide (forward or reverse strand, respectively) of the "gene" feature in the *gtf* annotation file. The distance of *Alu* elements to the TSS was analysed using the closest tool from BEDtools [93] and custom R scripts. Gene expression was calculated as the average expression (TPM) across all samples. The difference in distribution of gene expression was determined to be statistically significant using a two-sided Wilcoxon rank-sum test. A two-tailed Fisher's exact test was used to analyse the difference between the fraction of the set of all of expression-positive *Alus* and the fraction of the set of all unexpressed *Alu*s within each genomic range.

ChIP-seq analysis

ChIPs of TFIIIC (GTF3C2) and BDP1 were performed using procedures and antibodies described in [5] and references therein [94], [95]. All data was aligned to the hg38 human genome reference (GRCh38) and processed as in [5]. The average ChIP signal and heatmap profiles were visualized using the tools plotHeatmap from the deepTools package v3.5.1 [96].

TF binding motif analysis

The Analysis of Motif Enrichment (AME) from MEME suite [97] was used to analyze DNA sequences upstream of expression-positive Alus (200 bp). The Jaspar 2020 collection was used as motif database [98] and the DNA sequences of 200 bp upstream of 20000 random unexpressed Alus were used as a control.

Western blot

Cells infected with the adenovirus-based vectors were detached by scraping from a 60 mm plate and lysed in EBC buffer (50 mM Tris-Cl pH 8.0, 120 mM NaCl, 0.5% NP-40) with Roche cOmplete™ protease inhibitor cocktail. Samples were prepared in Laemmli buffer and heated for 5 minutes at 65°C. Protein extracts were resolved in an SDS-polyacrylamide gel and electrotransferred to a polyvinylidene difluoride (PVDF) membrane. Blocking was performed in 5% skim milk in TBS-Tween 0.1% for 10 minutes. Extracts were probed with antibodies against E1A (anti-e1a MAb M73) [99], Ku86 H-300 (sc-9034; Santa Cruz) and p400 (Thermo Fisher A300-541A) at manufacturer recommended dilutions for 1 hr at room temperature or O/N at 4°C. Membranes were washed 3 times with TBS-Tween 0.1% for 10 minutes at room temperature. Incubation with secondary antibodies was performed using anti-mouse or anti-rabbit IgG antibodies (Bio-Rad) in TBS-Tween 0.1% buffer with 5% skim milk for 1 hr at room temperature. Membranes were washed three times with TBS-Tween 0.1% for 10 minutes at room temperature before visualization with the Pierce ECL Western Blotting substrate (Thermo Fisher).

siRNA p400 Knockdown

Dharmacon SMART ON-TARGETplus pool siRNA against human EP400 (L-021272-05-0005) and D-001810-10-05 ON-TARGETplus Non-Targeting Pool were used to carry out EP400 knockdown in IMR90 cells. Cells were seeded in the absence of antibiotics and culture for 16h prior to transfection with lipofectamine (Lipofectamine 3000, Invitrogen). siRNAs were used at 30 nM and cells were left in culture for 24 h in the presence of the siRNA, followed by adenoviral dl1500- and e1a_p400-b- and mock-infection for other 24h. The knockdown efficiency was

evaluated by RT-qPCR (EP400_forward: GTGTTCGAGCGGTCACTTCT, EP400_reverse: CTGAACTCCTGTGGCTTGGG) after total RNA extraction (Trizol) and reverse transcription of 500 ng of RNA to obtain cDNA. Knockdown efficiencies of p400 were around 50-60% depletion in different experiments.

Reverse Transcription and Real-Time PCR

RNA extracted from infected cells (500 ng) was reverse-transcribed using SuperScript III Reverse Transcriptase (Thermo Fisher) with random hexamer primers. The Real Time PCR reaction was performed using 20 ng of cDNA and the PowerUp SYBR Green Master Mix (Applied Biosystems), in a 20 µl final volume with 300 nM primer concentration. Runs were performed using a 7500 Real-Time PCR System (Applied Biosystems). Expression levels were normalized using U1 snRNA as internal control and the ΔΔCt method was used to evaluate expression relative to mock-infections (U1_forward: AGGGCGAGGCTTATCCATTG, U1_reverse: GCAGTCGAGTTTCCCACATTTG). Data are presented as an average of two replicates ± standard deviation. Expression levels of *Alus* (*Alu*Sp_qRT_fw: TCACTCTGTCGCTTCCCAGTTG, *Alu*Sp_qRT_rev: TTAAAACTCTGCTCTGTTCACGCAG, *Alu*Sc_qRT_fw: GCTAGAGTCAACCTCACCTGC, *Alu*Sc_qRT_rev: TTAAGCCGCCAGAACGAGC for siRNA knockdown experiments were calculated using ΔΔCt (siEP400-siCTRL) method for each infection and compared to respective non-siRNA infections relative to mock.

Results

Small e1a is sufficient to induce Alus transcription

To investigate whether the presence of small e1a alone can induce upregulation of *Alu*s transcription we infected G_1-arrested IMR90 with a Human Adenovirus C5 (HAdV-C5) mutant named *dl*1500, that, due the presence of a 9-bp deletion in the 5' splice site expresses only small e1a. As negative control we employed a mutant called *dl*312 which does not express E1A or any other viral proteins [100]. As an additional control we performed a mock-infection. The changes in *Alus* expression at single locus resolution was assessed through RNA-sequencing with a sequence depth of 60 million paired end (PE) reads, to overcome the low abundance of *Alu*s transcripts. To evaluate the effect of e1a on *Alu*s expression alone, without transcriptional changes due to other transcribed loci that may harbour embedded *Alus,* we considered only the *Alus* transcripts from independent Pol3 transcription units, excluding intragenic *Alu*s in sense orientation. On the 802431 *Alu*s considered, only 1805 were detected in at least one replicate (we named them "expression positive *Alus*" or ep*Alu*s), with an overall overlap between the duplicates in each condition >75% [Figure 1].

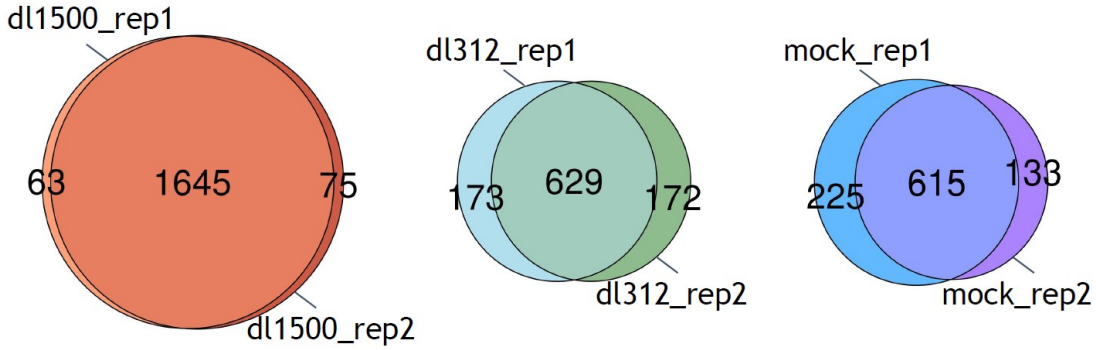

Figure 1. Detected *Alu*s transcripts for each experimental replicate. A total of 1783 *Alu*s were detected in *dl*1500 infected cells, 974 *Alu*s in *dl312* infected cells and 973 *Alu*s in mock-infected cells. The overall overlaps between replicates in each condition is >75%.

In response to *dl*1500 infection, and therefore the e1a expression, the *Alu* transcriptome expanded significantly, with the detection of 666 more *Alu*s (~37% of the total ep*Alu*s)

compared to the *dl*312 and mock infection, whereas only 22 *Alu*s were detected (~1%) [Figure 2].

Ep*Alu*s detected in mock- and *dl*312 infected cells presented a partially overlap, which may result from the interaction between cellular and viral factors, such as receptors and nucleic acids [101].

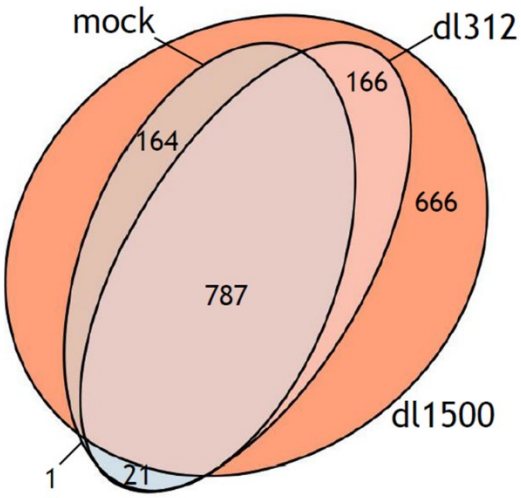

Figure 2. E1a expression causes an expansion of the *Alu* transcriptome. Venn-diagram showing the *Alu* transcriptome overlaps between dl1500, dl312 and mock-infected cells.

The global *Alu*s induction triggered by e1a is evident in [Figure 3], where we detected a ~4-fold increase in reads associated with *Alu*s after the *dl*1500 infection. Among the 953 *Alu*s that were expressed followed *dl*1500 or *dl*312 infected cells, 384 were upregulated (FDR < 0.05) whereas 1 *Alu*s was downregulated (FDR < 0.05). These data show how the viral protein e1a alone can derepress *Alu*s. In agreement with previous studies[11], [19], just the 0.12% of the intergenic/antisense *Alu*s were constitutively expressed in IMR90. Following the *dl*1500 infection, there was an increase to the 0.22%. Despite the increase in expression, induction by e1a remains a phenomenon that was limited to a few *Alu*s.

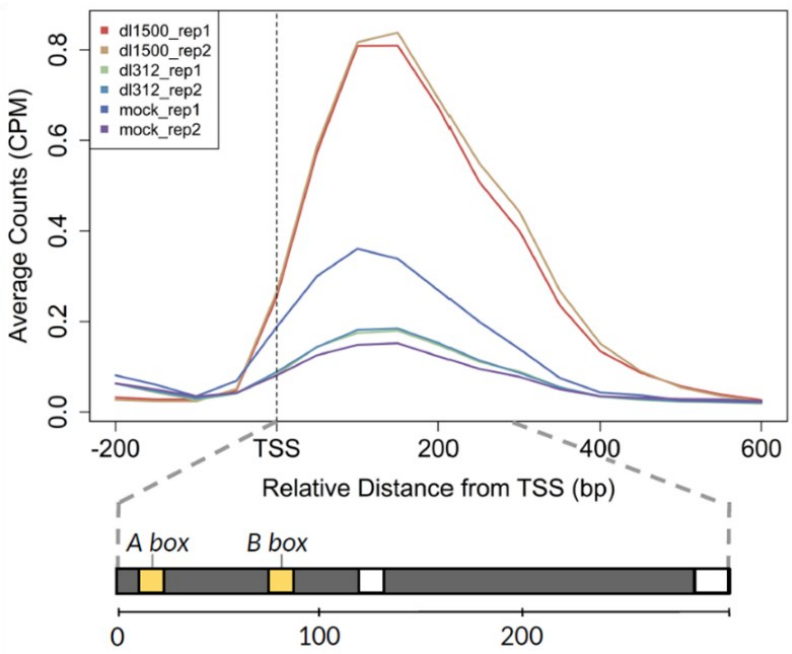

Figure 3. E1a induce an upregulation of *Alus* transcription. The graph shows the Alu-metagene profile of the average normalized number of reads (count per million, CPM) over the *Alu* region in the different conditions.

*Alu*S was the most represented subfamily between the ep*Alus*, whereas *Alu*J and *Alu*Y were underrepresented [Table 1]. This could be explained since *Alu*J, the ancient *Alus* subfamily, given the amount of mutation accumulated, is not able to be efficiently transcribed, while *Alu*Y, the youngest subfamily, is subjected to stricter repression due to its capacity for retrotransposition.

	AluJ	***AluS***	***AluY***	***No famiy***	***Tot***
Annotated	213441 (26.6%)	484941 (60.43%)	100903 (12.57%)	3146 (0.39%)	802431
Expressed (dl1500)	187 (10.49%)	1499 (84.07%)	96 (5.38%)	1 (0.05%)	1783
Subfamily enrichment[b]	0.39	1.39	0.43		
P-value[c]	$<2.2 \times 10^{-16}$	$<2.2 \times 10^{-16}$	$<2.2 \times 10^{-16}$		
Expressed (dl312)	155 (15.91%)	748 (76.8%)	70 (7.19%)	1 (0.1%)	974
Subfamily enrichment	0.60	1.27	0.43		
P-value	$<2.2 \times 10^{-16}$	$<2.2 \times 10^{-16}$	$<2.2 \times 10^{-16}$		
Expressed (Mock)	153 (15.72%)	750 (77.08%)	69 (7.09%)	1 (0.1%)	973
Subfamily enrichment	0.59	1.28	0.56		
P-value	$<2.2 \times 10^{-16}$	$<2.2 \times 10^{-16}$	$<2.2 \times 10^{-16}$		

Table 1 Subfamily distribution of ep*Alus*[a]. (a) For each *Alu* subfamily is reported the absolute number of intergenic/antisense *Alus* and their percentage. In the first row, the subfamily distribution of annotated *Alus* is reported. The other rows report the subfamily distribution of *Alus* expressed in *dl*1500-, *dl*312- and mock-infected IMR90 cells. **(b)** For each sample, the subfamily enrichment was calculated as the ratio between the percentage

of each subfamily out of all expressed *Alu*s, and the percentage of each subfamily out of all annotated *Alu*s**. (c)** The p-value for subfamily enrichment was calculated using a two-tailed Fisher's exact test.

TFIIIB is recruited at derepressed *Alu*s

*Alu*s transcription relies on a type II promoter of Pol3, so the recruitment of TFIIIC and TFIIIB is required. Therefore, given the strong *Alu*s induction we investigated whether e1a could affect the occupancy of TFIIIC and TFIIIB. To answer this, we studied the genome-wide localization of TFIIIC and TFIIIB in mock and *dl*1500 infected cell through ChIP-sequencing (ChIP-seq). The genome-wide occupancy of TFIIIC was tested using an antibody targeting TFIIIC110 (the second largest subunit), whereas the occupancy of TFIIIB was determined using an antibody against Bdp1 (one of the subunits of TFIIIB). We identified four different cluster of loci, divided according to the presence or absence of TFIIIB/TFIIIC following the *dl*1500 infection. Cluster 1 was characterized by an increase in Bdp1 enrichment that it was not associated with changes in TFIIIC occupancy, differently from what was observed in cluster 2 where there was also an increase in TFIIIC enrichment. These two cluster were enriched for the presence of tRNAs. Cluster 3 and cluster 4 were characterized by the presence of TFIIIC and absence of Bdp1, with a decrease in TFIIIC enrichment in cluster 4. These two clusters were strongly enriched in *Alu*s [Figure 4a and Figure 4b]. Interestingly, the Genomic Regions Enrichment of Annotation Tool (GREAT) analysis associated the cluster 4 loci to genes involved in embryonic lethality [Figure 4c]. The removal of TFIIIC from these loci in an e1a-dependent manner, could be part of the mechanism that led the transformation of cells caused by e1a.

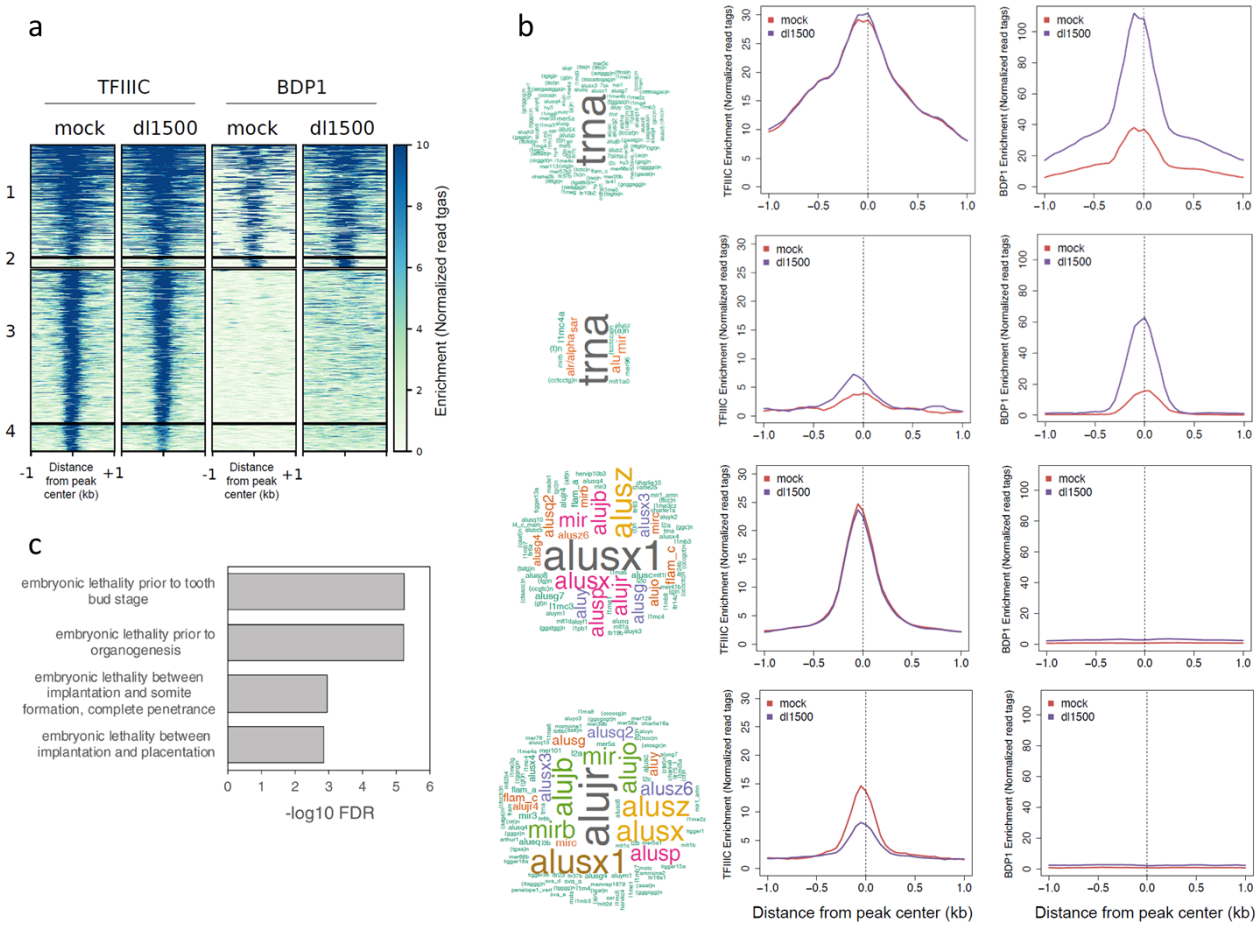

Figure 4. Genome-wide localization analysis of TFIIIC and TFIIIB in mock- and *dl*1500-infected cells. (a) Heatmap of TFIIIC and Bdp1 spanning ±1 kb across TFIIIC-bound sites in mock- and *dl*1500-infected cells. Clusters 1 to 4 were by combinatorial clustering of the two factors across all the regions bound. **(b)** On the left is reported the word cloud analysis of repetitive elements identified in the four clusters. The font size is proportional to the enrichment of the term. On the right are reported the results of the *sitepro* analysis [102] of TFIIIC and Bdp1 enrichment (normalized read tags) for each cluster. The enrichment is shown spanning ±1 kb. **(c)** Bar plots of the GREAT gene ontology (GO:Human Phenotypes) enrichment for TFIIIC-bound regions of cluster 4.

Focusing our analysis on the ep*Alu*s, we observed that e1a did not act on the recruitment of TFIIIC on ep*Alu* promoters, since TFIIIC was already present and these loci. Furthermore, since *Alu*s were sorted by expression level, a higher TFIIIC occupancy level was observed for those ep*Alu*s that were more highly expressed, both in the presence and absence of e1a [Figure 5a]. As expected, there was no TFIIIC enrichment when a random set of *Alu*s, was considered. Differently from what observed for TFIIIC, Bdp1 did not occupy ep*Alu*s in absence of e1a, whereas in *dl*1500 infected cell the occupation of Bdp1 on ep*Alu*s increased. This became more evident when we considered only the most expressed *Alu*s among the upregulated ones

[Figure 5a and Figure 5b]. In Figure 5c is reported the genome browser view of three *Alu*s associated with their expression profile and the ChIP-seq tracks for TFIIIC and Bdp1. Taken together, the data show that e1a led to the recruitment of Bdp1 (as a representative subunit of TFIIIB) on the upregulated *Alu*s, but did not act on TFIIIC occupancy, since this factor was already bound at the ep*Alu*s promoter, independently of the e1a presence.

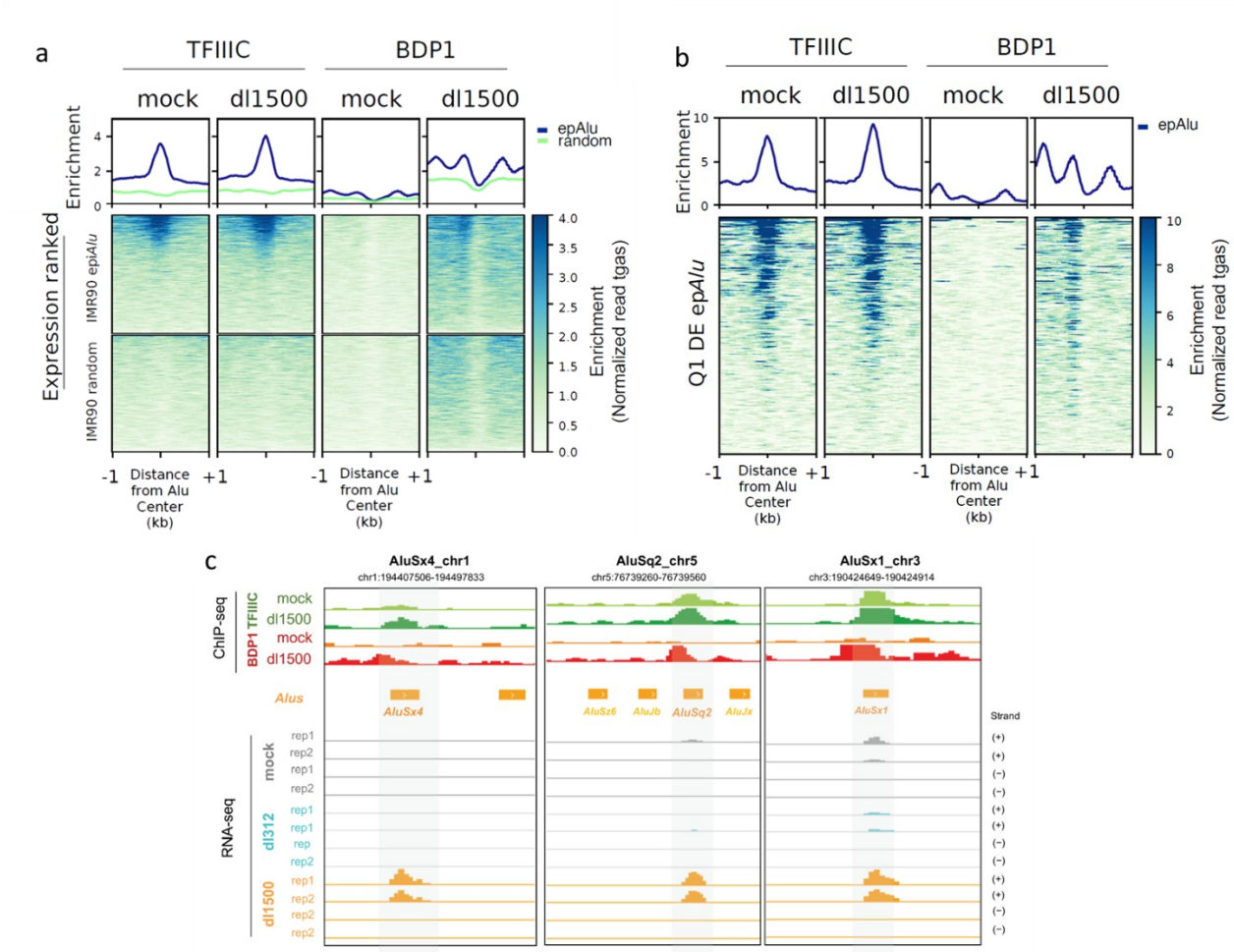

Figure 5. TFIIIC and Bdp1 localization across ep*Alu*s in uninfected and *dl*1500 infected cells. (a) In the upper part of the panel is represented the average ChIP-seq enrichment profile (normalized read tags) of TFIIIC and Bdp1 in either mock- or *dl*1500 infected cells, across the 1805 ep*Alu*s and across of a random set of *Alu*s with a similar size. Below are reported the heatmaps of TFIIIC and Bdp1 enrichment at the same *Alu*s, sorted in accordance with the degree of transcription in *dl*1500 infected cells (top, higher expression; bottom, lower expression). **(b)** Enrichments profiles (normalized read tags) of TFIIIC and Bdp1, in either mock or *dl*1500 infected cells, at differentially expressed *Alu*s whose expression in *dl*1500 infected cell falls in the first quartile (Q1 DE ep*Alu*s), sorted in accordance with the degree of transcription in *dl*1500 infected cells (top, higher expression; bottom, lower expression). **(c)** Genome browser views of TFIIIC and Bdp1 ChIP-seq and RNA-seq data of three highly expressed Alu elements. Yellow arrows represent the position and orientation of repetitive elements as evidenced by the RepeatMasker track. The chromosomal coordinates of each annotated *Alu* are shown above each view. Bigwig tracks are normalized for the library size.

ep*Alu*s have unique epigenetic features

*Alu*s generally show epigenetic features that repress their expression [103]. We investigated whether ep*Alu*s presented unique epigenetic signatures and how e1a influenced them. We took advantage of public datasets of MNase-seq [104] and ATAC-seq [105] performed on IMR90 to analyse the chromatin organization of ep*Alu*s. Through DANPOS [106] we analysed the MNase-seq data to determine the nucleosome enrichment across the *Alu*s. As shown in Figure 6a, ep*Alu*s and a random set of *Alu*s are occupied by two nucleosomes. Despite the similar occupancy across the *Alu*s, ep*Alu*s presented a more open chromatin upstream of the TSS compared to a random set of *Alu*s [Figure 6b].

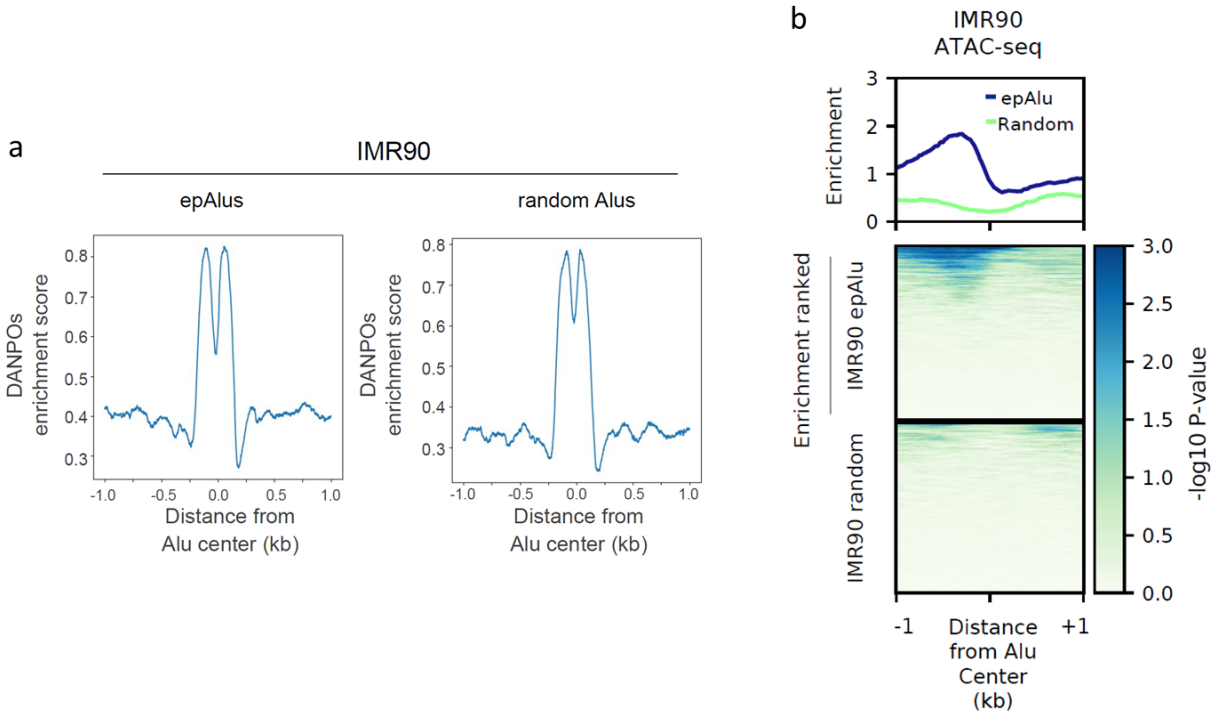

Figure 6. Chromatin accessibility around *Alu*s. (a) Nucleosome density has mapped by MNase digestion across ep*Alu*s and random *Alu*s in IMR90 cells using the DANPOS score [106]. The plot has been produced using NucMap [104]**. (b)** Heatmap of ATAC-seq signal around across the 1805 ep*Alu*s and across of a random set of *Alu*s with a similar size. Ranking is according to enrichment of ATAC-seq reported as -log10 of the Poisson p-value.

Ep*Alu*s have histone modification patterns that resemble enhancers [19]. We performed ChIP-seq for H3K18ac, H3K27ac and H3K4me1, features of active enhancers and H3K9ac feature of active transcription, to learn whether these marks changed following *dl*1500 infection [Figure 7]. In mock-infected cells, ep*Alu*s, compared to a random set of *Alu*s, presented a strong enrichment in H3K27ac and a limited presence of H3K9ac and H3K18ac. H3K4me1 was neither

enriched in ep*Alu*s nor in a random set. The *dl*1500 infection triggered changes in the H3K27ac and H3K4me1 profile in ep*Alu*s, the former being depleted while the latter increased upstream of the TSS and in the body of the ep*Alu*s. H3K18ac and H3K9ac changes were less evident and not specific. These data suggest that ep*Alu*s had enhancer-like characteristics and e1a could act on their epigenetic profiles, depleting H3K27ac and causing the deposition of H3K4me1 setting ep*Alu*s in a state that resembles that of poised enhancers [35] .

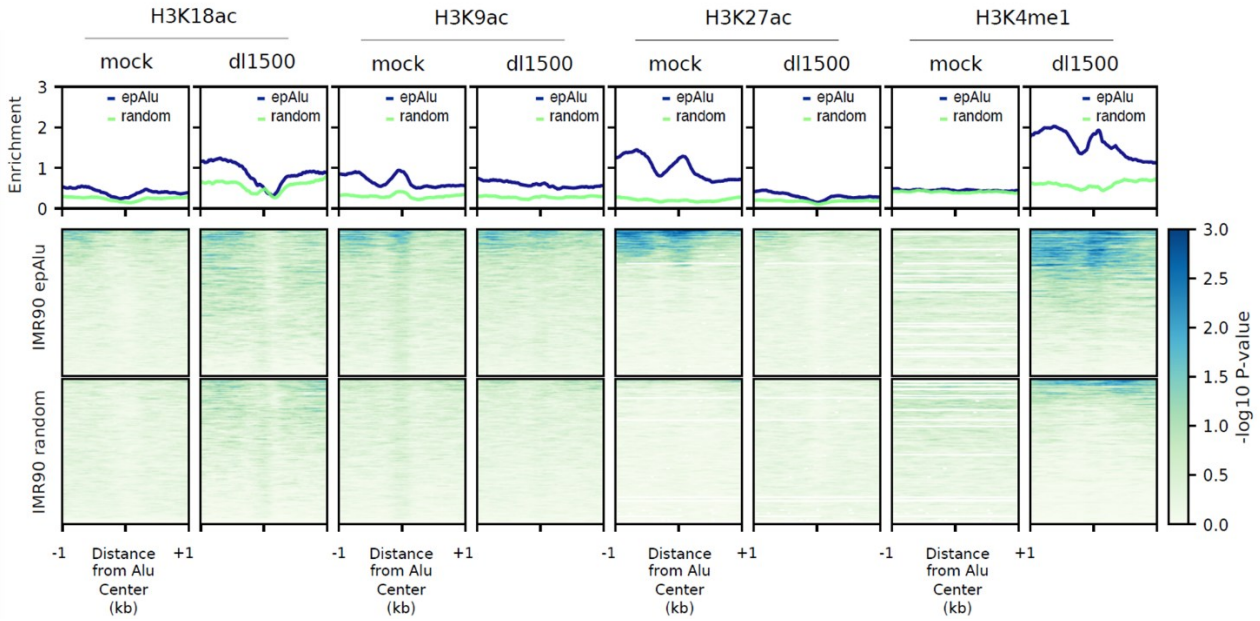

Figure 7. Histone modification enrichment profiles of ep*Alu*s in uninfected and *dl*1500 infected cells. Shown in the upper graphs are the average ChIP-seq enrichment ($-\log_{10}$ of the Poisson *P*-value) profiles of (from left to right) H3K18ac, H3K9ac, H3K27ac and H3K4me1 across the 1805 ep*Alu*s in either mock- or *dl*1500-infected IMR90 cells. Reported below the plots are the heatmaps of the same histone modification enrichments, with ep*Alus* and random *Alus* ranked according to enrichment expressed as $-\log_{10}$ of the Poisson *P*-value.

p300 and Rb are two known e1a interactors important for the e1a-mediated reprogramming of host epigenome due to their genome-wide redistribution upon *dl1500* infection [67]. We therefore asked whether p300 and Rb could potentially being implicated in *Alu*s induction by e1a. To answer this, we implemented ChIP-seq of p300 and Rb in dl1500-infected IMR90 (compared to mock) to verify their possible occupancy at ep*Alu* loci [Figure 8]. We found that p300 and Rb occupied the same region upstream of the *Alu* and, as observed for TFIIIC, e1a did not affect their occupancy at these loci. Furthermore, the enrichment of these factor in mock and *dl*1500 infected cell was higher for those *Alu*s which were more induced by e1a. Importantly, the p300 and Rb presence was characteristic of ep*Alu*s, since when a random set

of *Alu*s was considered, we found no enrichment. These data showed that the behaviour of p300 and Rb on ep*Alu*s was comparable to what observe for TFIIIC, being present independently of e1a expression. Nevertheless, they marked the *Alu* loci that will be upregulated by e1a.

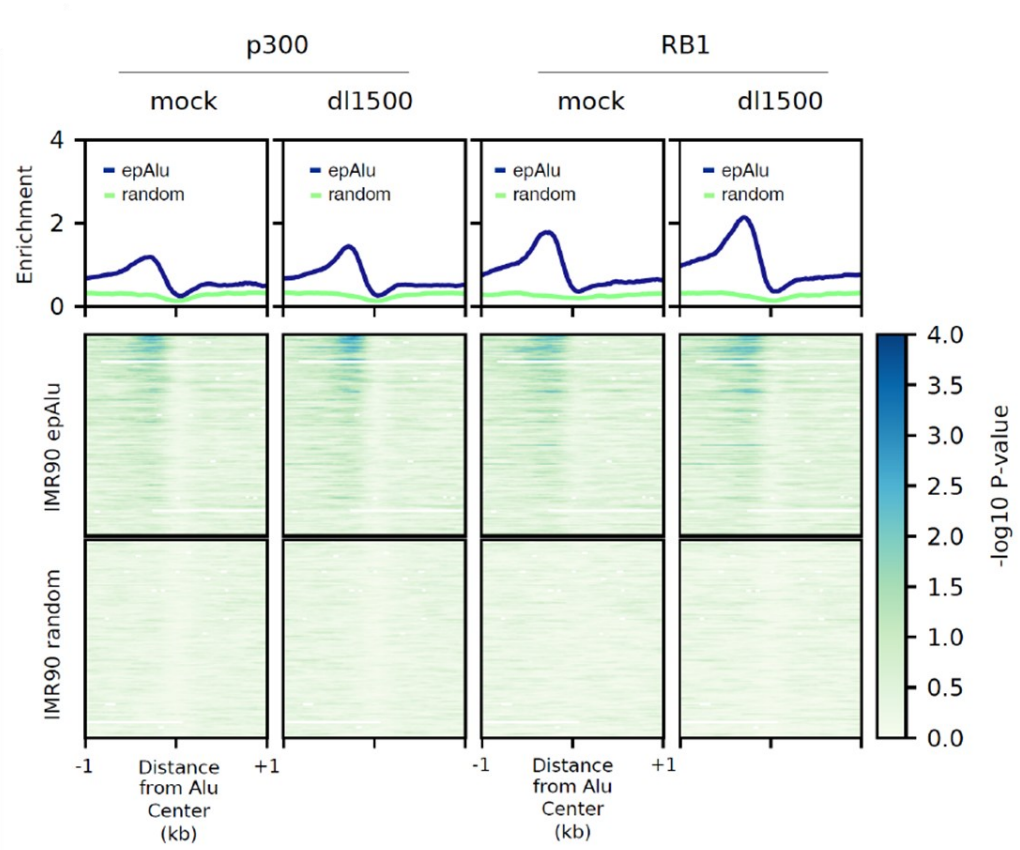

Figure 8. Chromatin regulator enrichment profiles of ep*Alu*s in uninfected and *dl*1500 infected cells. Shown in the upper part of the panel are the average ChIP-seq enrichment profiles (–$\log_{10}$ of the Poisson *P*-value) of p300 (left) and RB1 (right) in either mock- or *dl*1500-infected IMR90 cells across the 1805 ep*Alu*s and across random *Alu*s. Reported below the plots are heatmaps of p300 and RB1 association to the same *Alu*s, sorted according to their expression level in *dl*1500-infected cells (top, high expression; bottom, low expression).

epAlus loci are bound by enhancer-associated factors

Since the expressed *Alu*s have enhancer-like characteristics and harbour several transcription factors binding sites [11], [19], we implemented *in silico* prediction of possible transcription factors binding signature on ep*Alu*s. We performed Analysis of Motif Enrichment (AME) [107] on the 200 bp upstream of ep*Alu*s, unveiled that AP-1 was the most represented motif. Other binding motifs identified were those for the TEAD proteins family, NFIX, CTCF, ZBTB7A, ETV6,

BLHAHA15 and CEPBE [Figure 9a]. AP-1, CTCF and CEBPB were already known to occupy the loci of expressed *Alu*s [11], [19]. The analysis of published ChIP-seq through the Cistrome Toolkit [108] allowed us to search proteins that were found enriched on the ep*Alu*s identified in this study in other ChIP-seq experiments [Figure 9b]. As illustrate in Figure 9b, higher enrichment was observed for components of the Pol3 transcription system (POLR3D, POLR3A, GTF3C2, TBP, POLR2H and GTF3C5). In addition, our analysis also showed enrichment for components of the ChAHP complex (ADNP, CHD4 and CBX3) recently been shown to bind *Alu*s that were bound by TFIIIC following serum-starvation in T47D and known to counteract the CTCF binding in many B2 SINEs. Another transcription factor motif enriched in ep*Alu*s was RCOR1, a corepressor interacting with ZMYM2 [109]. ZMYM2 is a known interactor of TFIIIC as well as ADNP at SINEs loci, which controls their transcription [110]. We also identified transcription factors expected to be recruited at the predicted binding motifs, such as YAP1 and BRD4, that are interactor of TEAD [111]; the AP-1 subunits JUN and FOSL1; CEBPB whose binding site is the same of CEBPE. The data generated by motif analysis showed a wide range of putative transcription factors that could potentially regulate the expression of ep*Alu*s.

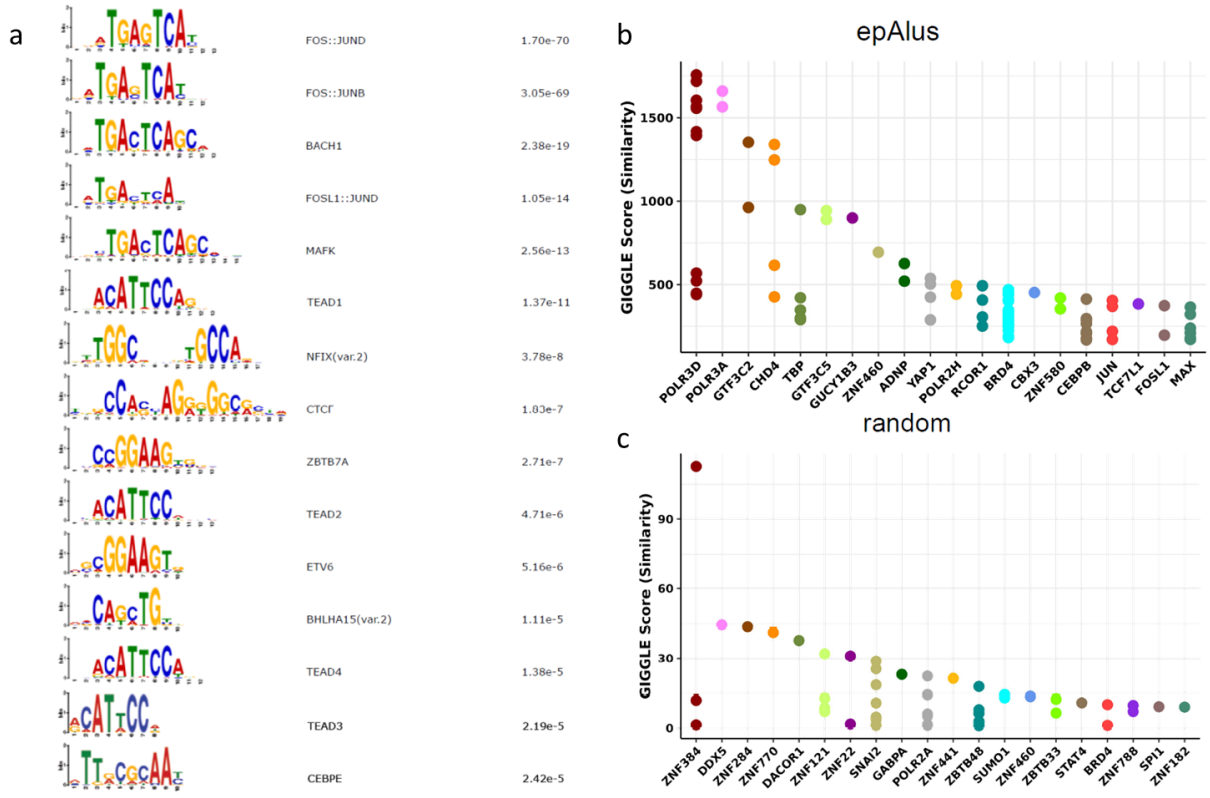

Figure 9. *In silico* prediction of transcription factors bound to ep*Alu*s. (a) TF binding motifs enriched in the 200 bp upstream of the TSS of ep*Alu*s with (right column) their corresponding P-values adjusted for multiple testing with Bonferroni correction**. (b)** Cistrome ToolKit analysis. GIGGLE score is reported [112] for ep*Alu*s and random *Alu* sets. Giggle score is calculated by using the genome coordinates of the two sets of Alus to retrieve which factors bind those intervals among all curated experiment in the Cistrome database[108] **.**

Given the results of the AME analysis, we focused our attention on the regulation of these loci in IMR90 cells, by analyzing published ChIP-seq data for YAP1, FOS, BRD4 and CEBPB [113], [114], [115], [116], [117]. As shown in Figure 10a, these factors resulted enriched upstream of the ep*Alu*s, whereas no enrichment was found in a random set of *Alu*s. We did not have information about the behaviour of these factor in *dl*1500 infected cells, but as unveiled previously (for p300 and Rb) YAP1, FOS, BRD4 and CEBPB could be bound to *Alu*s that had a higher transcriptional potential.

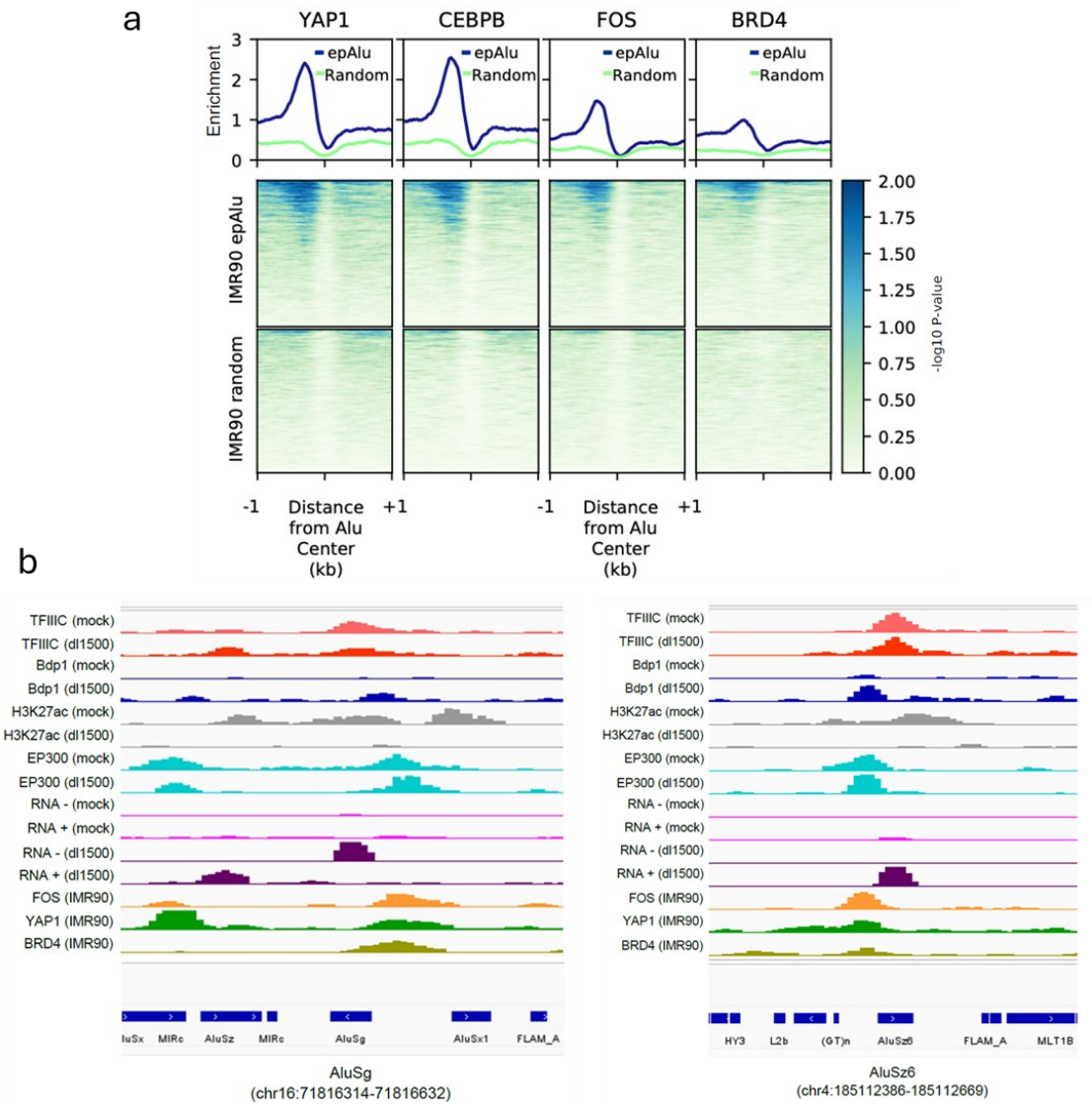

Figure 10. Enhancer associated factors enrichment profiles of ep*Alu*s in IMR90 cells. (a) Heatmaps of ChIP-seq of YAP1, CEBPB, FOS and BRD4 at the 1805 ep*Alu*s and at random *Alu*s based on data from [113], [114], [115], [116], [117]. Ranking is according to enrichment of YAP1 reported as $-\log_{10}$ of the Poisson *P*-value**. (b**) Genome browser views of TFIIIC, Bdp1, H3K27ac and p300 ChIP-seq data and representative RNA-seq data of two differentially expressed *Alu*s elements obtained in this study either in mock- or *dl*1500 infected cells, along with the views **of** publicly available ChIP-seq data for FOS, YAP1 and BRD4 in IMR90 cells. The chromosomal coordinates of each annotated *Alu*s are shown under each view in correspondence of the position of the e1a-induced *Alu*s.

The putative ep*Alu*s transcription factors identified previously are known to bind enhancer regions [117], [118], [119], [120]. To investigate more on the enhancer-like epigenetic features of ep*Alu*s, we focused our attention on the histone variant H2A.Z. The deposition of this histone variant depends on p400, an e1a interactor [73]. To overcome the lack of H2A.Z and p400 ChIP-seq data in IMR90 cells, we took advantage of published ChIP-seq collected in the K562 cell line [114], [116]. We found 3764 ep*Alu*s in K562 and these showed enriched for both H2A.Z and p400, while a random set *Alu* showed no enrichment [Figure 11b]. The enrichment was still present when we considered the ep*Alu*s detected in IMR90 [Figure 11a] suggesting the existence of a set of *Alu*s commonly expressed in all cell lines. Taken together these data showed that ep*Alu*s were likely enriched by enhancer-associated factors.

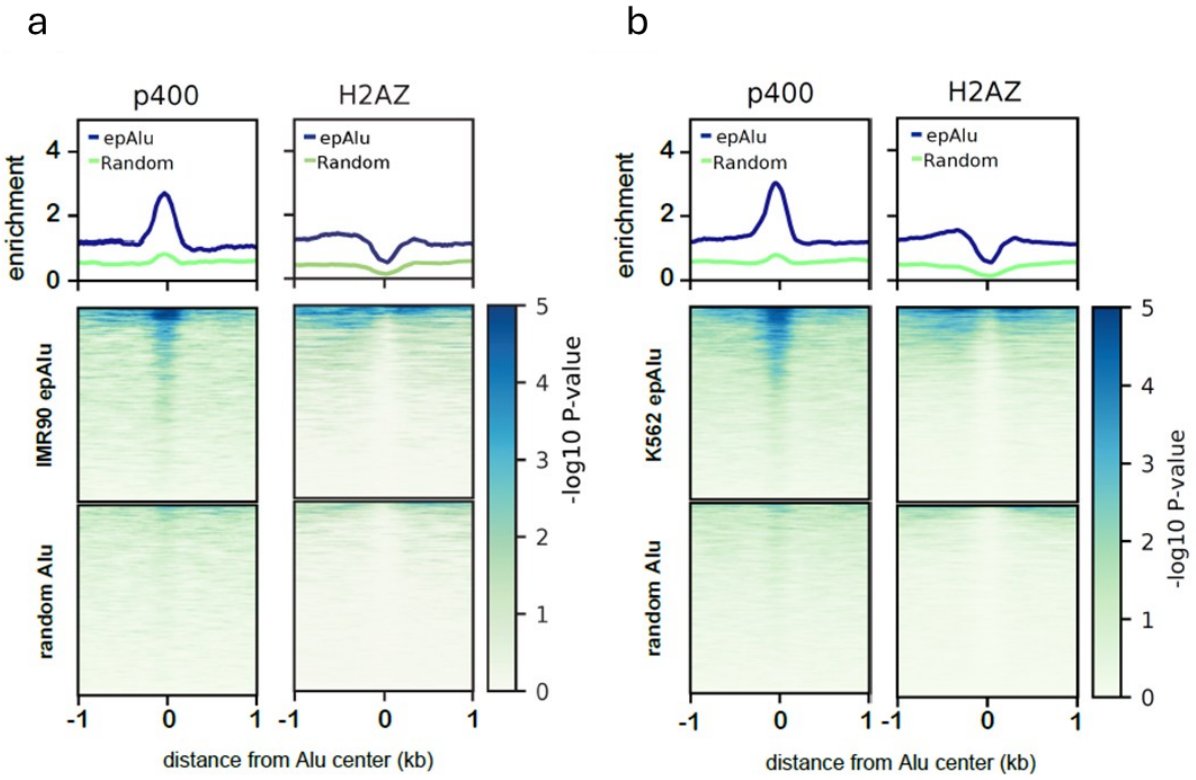

Figure 11. Enrichment of EP400 and H2A.Z at ep*Alus*. (a) Heatmaps of ChIP-seq enrichment of EP400 (left) and the H2AZ histone variant (right) in K562 cell[114], [116] across either the 1805 IMR90 ep*Alus* or random *Alus*. Ranking is according to enrichment of EP400 in K562 cells reported as -$\log_{10}$ of the Poisson *P*-value. **(b)** Heatmaps of ChIP-seq enrichment of EP400 (left) and the H2A.Z histone variant (right) in K562 cells across either the 3764 *Alus* detected as expressed in K562 cells or random *Alus*. Ranking is according to enrichment of p400 in K562 cells reported as -$\log_{10}$ of the Poisson *P*-value.

The interaction e1a-P400 is necessary for *Alus* upregulation

Our previous analysis showed that potentially many transcription factors could be bound to ep*Alus* but failed to explain the direct molecular mechanisms that drive the induction of *Alus* transcription upon adenovirus infection. Since we observed that p300, Rb and p400 occupied ep*Alus*, we asked if any of these factors could be necessary for ep*Alus* transcription. To answer this question, we infected IMR90 cells with *dl*1500 (wt e1a) and mutants viruses harbouring e1a mutations that specifically impaired its ability to interact with p300/CBP (e1a_p300_b-), Rb (e1a_Rb_b-) or p400 (e1a_p400_b-)[67] and we performed RNA-seq to define the *Alus* transcriptome. Previous studies shown that these e1a mutants have different stability or lower expression levels [67]. To obtain a comparable amount of e1a expressed across all the samples, we first identified the proper MOI for the different mutants by evaluating e1a protein levels through Western blot and e1a mRNA expression in our RNA-seq [Figure 12a and Figure 12b]. We found that the mutant e1a_Rb_b- was less prone to be transcribed and this was reflected in the less amount of e1a detected in the protein extract [Figure 12b]. Despite this was a not optimal condition, we proceeded with the RNA-seq, since e1a_Rb_b- has been shown to repress host genes more than the wild-type [67].

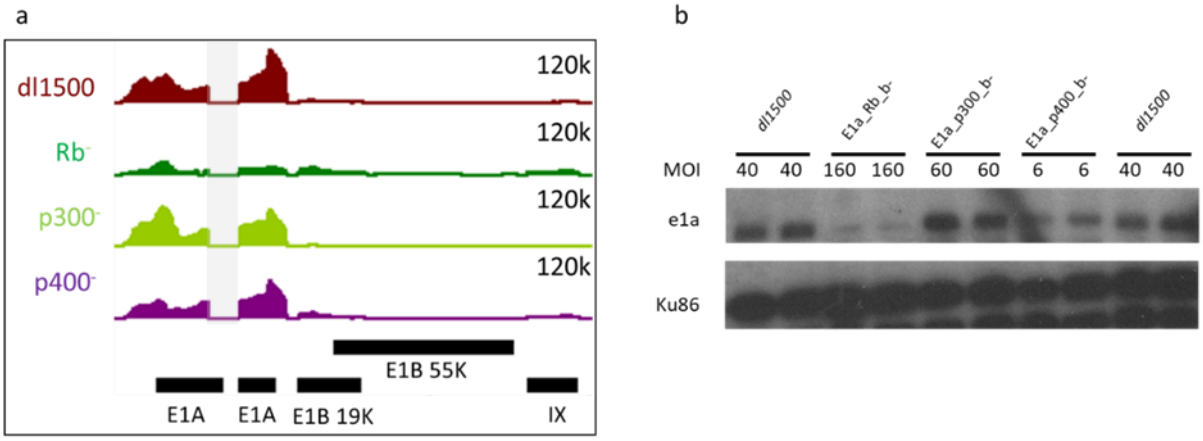

Figure 12. Expression of e1a in IMR90 cell infected with *dl*1500 and viruses expressing e1a mutants. (a) Expression profiles, shown as Integrated Genome Browser views, of the E1A gene in IMR90 infected with dl1500 (e1a-wt) or its interaction mutants. In grey is highlighted the CR3 coding region, absent in small e1a transcript. **(b**) Western blot showing protein levels of wild type e1a and e1a interaction mutants obtained from virus infections of IMR90 cells at the MOIs used for the experiment in Figure 13. Ku86 was used as a loading control.

Two biological replicates were used for the RNA-seq analysis. The identification of *Alu*s differentially expressed in the different infections was conducted comparing the four conditions with a mock-infected sample and reporting the result as a heatmap [Figure 13a]. We observed a different effect of e1a_p400_b- compared to e1a_p300_b- and e1a_Rb_b- despite this last mutant showing an increase in the total number of *Alu*s up-regulated [Figure 13a]. This could be explained by Rb having a negative modulatory effect on the *Alu*s transcription induced by e1a. In contrast, the infections with the e1a_p400_b- mutant caused a striking reduction of *Alu*s transcription, less evident with the e1a_p300_b- mutant [Figure 13a] in line with what has been reported for p300/CBP having a stimulatory effect on the transcription of Pol3 genes [121]. The effect of p400 was very surprising as 90% of the e1a-induced *Alu*s were no longer activated, whereas transcription of other Pol3 genes such as 7SLRNA, 7SKRNA, U6 snRNA, RNAse P RNA, $tRNA^{Gly}$, were not affected by this mutant [Figure 13c]. These results were further validated by RT-qPCR on two different ep*Alu*s [Figure 13b]. Taken together these results clearly showed that the *Alu*s upregulation depended on the interaction between e1a and p400.

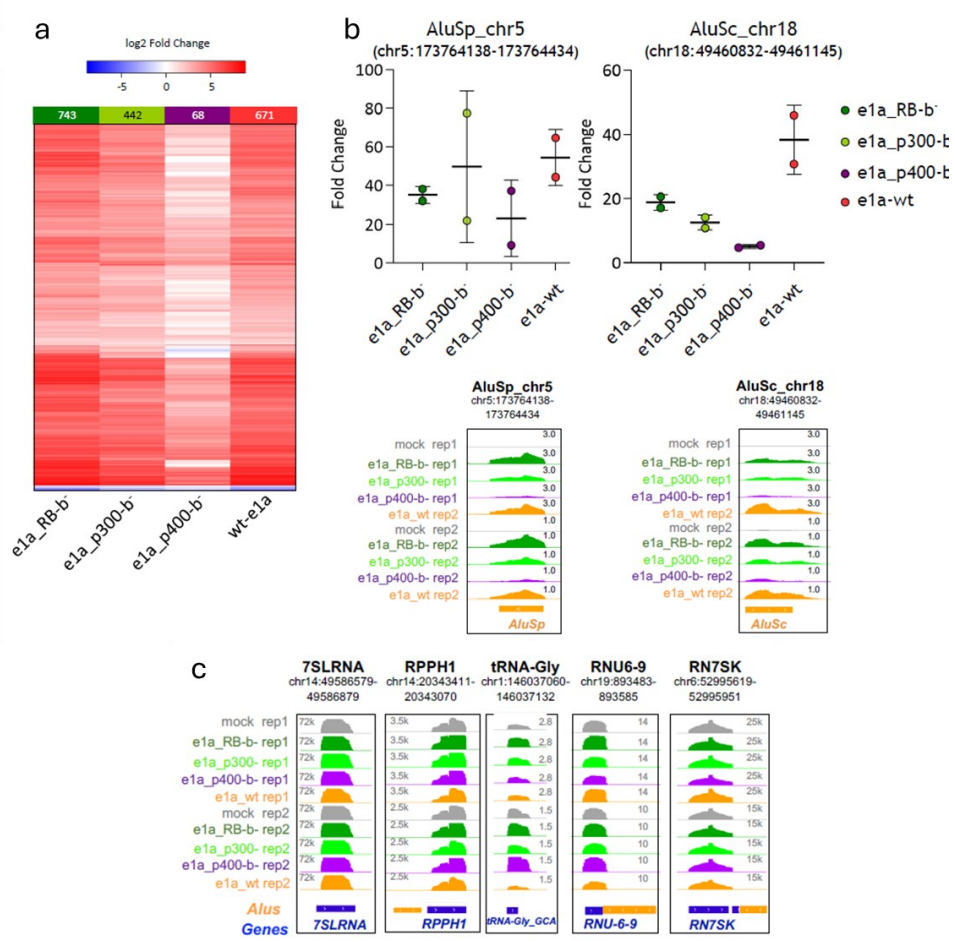

Figure 13. Dependence of *Alu*s transcription to e1a-host proteins interaction. (a) Heatmap showing increased (red) or decreased (blue) expression of *Alu*s triggered by wt e1a or e1a mutants defective in interaction with RB (e1a_RB-b$^-$), p300 (e1a_p300-b$^-$) or p400 (e1a_p400-b$^-$), as compared to mock-infected cells. Boxed above each heatmap are the numbers of differentially expressed *Alu*s (log_2 fold-change ≥ 0.5 or ≤ –0.5 and an adjusted *P*-value < 0.05). The experiment was performed in two biological replicates. **(b)** Expression levels of two individual *Alu*s as measured by RT-qPCR (upper graphs) and RNA-seq (lower views). Fold changes estimated by RT-qPCR are relative to the expression in mock-infected cells, after normalization to U1 snRNA gene expression. Primers were chosen to target the unique sequence of *Alu*s within the 3' trailer region. RT-qPCR data relative to each independent experiment are represented as dots. Indicated by horizontal bars are the means ± standard deviation between the replicates. RNA-seq data (lower subpanels) are presented as genome browser views of the same *Alu*s analysed in the upper plots. Orange boxes represent the orientation of repetitive elements as evidenced by the RepeatMasker track. The chromosomal coordinates of each annotated *Alu* are shown in the upper part of each subpanel. Bigwig tracks are normalized per CPM. **(c)** Genome browser views of the expression of RN7SL1, RPPH1, RNU6-9 and RN7SK genes. Expression profiles are based on RNA-seq analysis of IMR90 cells infected as indicated on the left**.**

Based on our findings, Bdp1 was recruited on differentially expressed *Alu*s and ep*Alu*s in an e1a-dependent manner [Figure 5] and the global induction of *Alu*s transcription depended on the e1a-p400 interaction. We therefore asked whether the Bdp1 recruitment could also depend on the interaction between e1a and p400. To answer this, we performed Bdp1 ChIP-

seq in dl312-, dl1500- and e1a_p400_b- infected cell [Figure 14a and Figure 14b]. We found that infection with e1a_p400_b- revealed a decrease Bdp1 recruitment on differentially expressed *Alu*s and epAlus compared to dl1500 virus. To understand whether the mechanism was dependent on p400 displacement by e1a (to allow the Bdp1 recruitment) or if p400 was required to activate transcription, we employed siRNA knock down of p400 prior infection with dl1500, e1a_p400_b- viruses compared to mock and evaluated the expression of same two *Alu*s previously validated by RT-qPCR [Figure 13b]. Knock down of p400 was confirmed by RT-qPCR and by Western Blot [Figure 14c]. p400 depletion did not triggered *Alu*s upregulation in mock and e1a_p400_b- infected cells, providing evidence that p400 was necessary for *Alu*s transcription induced by e1a as *Alu*s induction is abolished in dl1500 infected cell that were depleted of p400 [Figure 14d]. Taken together these data showed that e1a-dependent *Alu*s induction relied on the presence of p400 and its availability to interact with e1a.

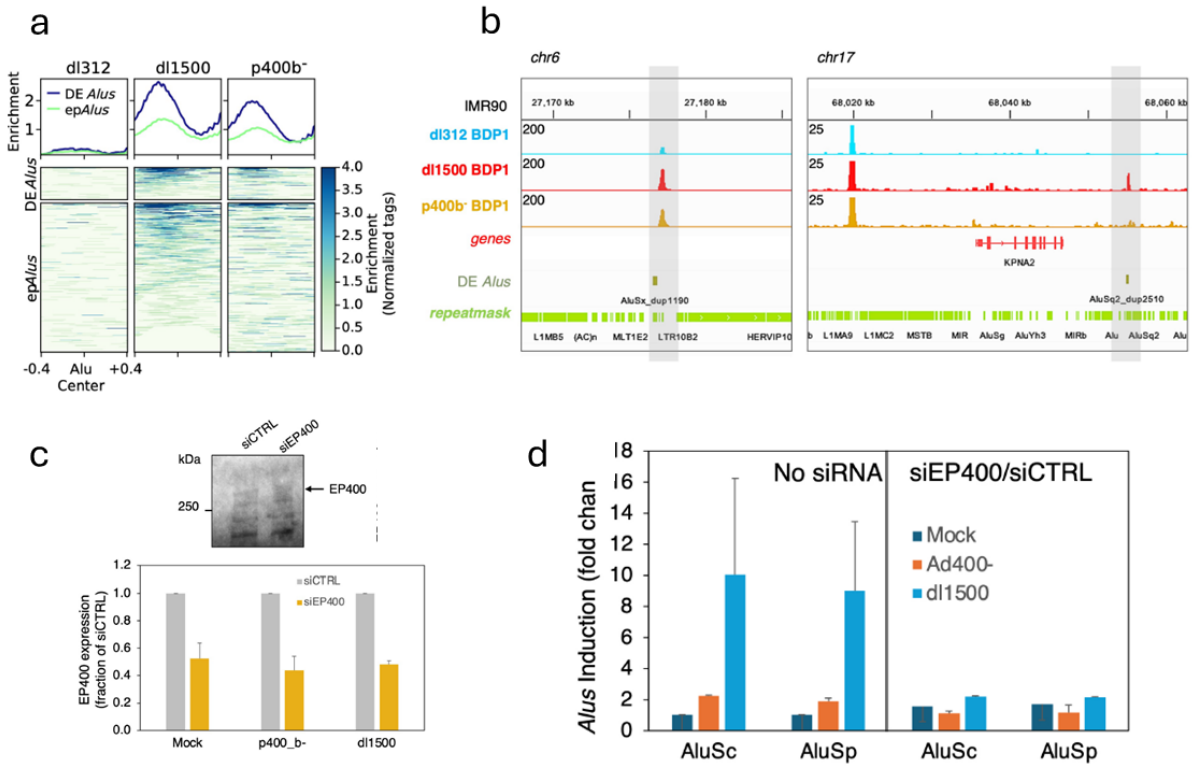

Figure 14. Dependence of *Alu* transcription to p400 chromatin remodeller. (a) Heatmap and enrichment profiles (normalized read tags) of Bdp1 ChIP-seq occupancy at differentially expressed *Alu*s (DE ep*Alu*s) and ep*Alu*s in IMR90 infected with *dl*312, *dl*1500 and p400-b⁻ viruses. **(b)** Genome browser views Bdp1 ChIP-seq data of two highly *dl*1500-induced *Alu*s as evidenced by the RepeatMasker track. The chromosomal coordinates of each annotated *Alu* are shown above each view. Bigwig tracks are normalized for the library size**. (c)** In the upper

panel is shown the western blot of p400 KD in IMR90 fibroblasts treated with scramble siCTRL or siRNA smart pool against EP400. Arrow indicates the band corresponding to EP400. In the panel below is represented the RT-qPCR measuring the expression of EP400 transcript upon treatment with siRNA smart pool against EP400. Levels of EP400 in IMR90 fibroblasts treated with scramble siCTRL are set to one**. (d)** RT-qPCR for measuring expression of two ep*Alu* loci (the same as in Figure 13b) comparing mock-, e1a_p400-b^{-} and *dl*1500-infection in conditions of absence of silencing RNA (non-siRNA) or presence of siRNA against p400 (siEP400) compared to a scramble set of siRNA control (siCTRL). Standard error bars are indicated, as result of two biological replicates.

Discussion

Adenoviral infection has been known to cause an increase in *Alu*s transcription [55]. The data we collected so far has shown that the adenovirus-5 small e1a was sufficient to induce the transcription of many *Alu*s displaying enhancer features associated with YAP/TAZ and AP-1 in IMR90 fibroblasts, in virtue of its interaction with the chromatin remodeller p400. E1a reprogrammed these loci to a state reminiscent of that of poised enhancers, displaying lack of H3K27ac but enrichment of H3K4me1. This effect was framed in a context in which e1a led a major reorganization of the host epigenome, to set up the best environment for viral replication [67], [68] . The e1a interactions with p300/CBP and Rb are the most studied and are responsible for the reprogramming of the transcription profile of the host cells [67]. Despite the interaction with p400 is known to promote the transformation of host cells [73], and to stabilize Myc through the formation of p400-Myc complexes on Myc target genes [83], it remains poorly studied. To date it is not known if Myc is also involved in *Alu*s induction, but N-Myc, a member of Myc proteins, has been reported to interact with TFIIIC and colocalize with it at several loci [122] whereas c-Myc, together with Gcn5 and TRAPP, activates Pol3 transcription of tRNAs and 5SRNA genes [16]. TRAPP is a p400 interactor in the Tip60/p400 complex and an e1a interactor [66]. Considering this, Myc proteins could potentially participate in *Alu*s induction. P400 is a chromatin remodeller that is involved in deposition of the histone variant H2A.Z, whose enrichment we observed on those *Alu*s that were actively transcribed in IMR90 cells agrees with a previous study [19]. E1a-p400 interaction could also promoting the H3K4me-dependent recruitment of the Tip60/p400 [76]. Our *in silico* analysis of ep*Alu*s also identified another putative chromatin remodeller, CHD4, a component of the ChAHP complex, a known interactor of TFIIIC and enriched at SINE loci [110]. *Alu*s are thought to be a source for the *de novo* birth of enhancer, and this is reflected in in their epigenetic features [19], [35]. Our results agree with the enhancer-like nature of *Alu*s previously described [19]. Indeed, e1a led ep*Alu*s to acquire a state reminiscent of that of a poised enhancer. The presence of p300 upstream of ep*Alu*s was also consistent with the enhancer-like features of these elements. The depletion of H3K27ac, suggested that at these loci p300 HAT activity was inhibited by e1a, as observed in other genomic regions. Poised enhancers are typically associated with those genes involved in cellular differentiation [123] and ep*Alu*s

epigenetic remodelling could be linked to cellular dedifferentiation caused by e1a [124]. YAP/TAZ coactivators have been reported to be involved in this process, since e1a causes a genome-wide loss of these factors at enhancer that are involved in differentiation [124]. In our analysis we found YAP/TAZ associated with ep*Alu*s loci in uninfected IMR90 cell, probably through the interaction with the TEADs family of transcription factors, whose binding motifs was enriched upstream of the ep*Alu*s. It was also reported that p400, p300 and BRD4 (that we also found to be enriched in uninfected IMR90 cells at ep*Alu*s) interact with YAP/TAZ coactivators. We had not investigated the role and the possible redistribution upon *dl*1500 infection of YAP/TAZ but is reasonable to hypothesize their involvement in the enhancer manipulation directed by small e1a. Possibly related to this unexplored mechanism are the interaction between p400 and YAP/TAZ and the enrichment of BRD4 at the ep*Alu*s, which is reported to mediate YAP/TAZ-dependent transcriptional activation [111]. In our analysis we found also an enrichment of the binding motifs, upstream at ep*Alu*s, as well as enrichment in ChIP-seq data, for the enhancer-associated transcription factors CEBPB and AP-1, in agreement with what observed in previous study [19]. CEBPB and AP-1 could exploit a role in transcription of ep*Alu*s through the recruitment of p300/CBP however the involvement and modulation of AP-1 at these loci requires further investigation since large E1A has been reported to abrogate the transcriptional activation activity of c-*fos* competing for p300/CBP binding [125]. YAP/TAZ and AP-1 have been observed to co-occupy enhancer regions associated with cancer progression [111]. These results suggest that *Alu*s upregulation could be part of a program to control enhancer functions of the host cell by e1a. Another point to be clarify is which isoform of transcription factors were involved in the *Alu*s induction. CEBP family members have binding sites almost identical, therefore ChIP studies are necessary to identify which isoforms were involved in *Alu*s transcription. Notable, we observed CEBPB enrichment at IMR90 ep*Alu*s. CEBPB enrichment at ep*Alu*s was also reported in other cell lines [126]. Another interesting point is the observation that the transcription factors binding sites were upstream of the *Alu*s body. *Alu*s promoter could synergize with the upstream flanking region to sustain the *Alu*s transcription [11]: this observation could in part explain the small number of transcribed *Alu*s, since most of them (even the *Alu*s that are found in an epigenetic context that makes the chromatin accessible) were not associated with these upstream flanking regions.

Previous studies showed that the *Alu*s transcripts accumulation might contribute to the remodelling of the cell transcriptome [127], [128]. However, this phenomenon has been observed in a context of artificial *Alu*s over-expression, that led to the production of higher number of transcripts compared to the one triggered by the e1a. Thus, is not possible to precisely determine if the *Alu*s upregulation has a role in cell transformation.

The e1a ability to increase the transcriptional activity of TFIIIC and TFIIIB was already partially associated with its interaction with Rb [129], [130]. Despite our ChIP-seq data showed an e1a-dependent recruitment of Bdp1 (TFIIIB), we observed that the abolishment of e1a-Rb interaction caused an increase of *Alu*s transcription. ChIP-seq against Rb performed in IMR90 cells, showed that the tumour suppressor was stably bound to many tRNA genes [131], suggesting its plausible role in steps following TFIIIB recruitment.

The most relevant genome-wide TFIIIC redistribution following *dl*1500 infection occurred in a subset of regions enriched in *Alu*s that showed decreased TFIIIC levels upon e1a-expression. These *Alu*s did not overlap with ep*Alu*s, thus the TFIIIC modulation at these loci did not affect their transcription. Notably, these genomic regions were associated with genes involved in embryonic development and therefore it is possible that the depletion of TFIIIC from these regions might contribute to cell de-differentiation induced by e1a [132]. The possible role of TFIIIC in differentiation pathways has also been reported by several CRISPR-Cas9 screening performed in human embryonic stem cells (hESCs) [133], [134], [135]. Based on these data we hypothesized that TFIIIC and *Alus* might play a role in hESCs maintenance and differentiation, and this question will be addressed in the following chapter of my thesis.

Chapter2

Introduction

*Alu*s involvement in development and stemness

As pointed out above, expressed *Alu*s are tissue- and cell- specific and have epigenetic marks typically associated with enhancer regions. Human embryonic stem cells (hESCs) have a specific pattern of expressed *Alu*s, that are enriched on active enhancers bounded by transcription factors associated with stemness, such as OCT4 and NANOG [19]. *Alu*s transcription was also associated with the formation of chromatin loops [87], [136]. *Alu*s transcribed as enhancer RNA (eRNA) and promoter upstream antisense RNAs (uaRNA) foster the specific pairing between enhancer and promoter through the formation of enhancer-promoters RNA interaction (EPRI) in hESCs and other cell lines [87]. *Alu*s-derived eRNAs can also participate in the formation of R-loops (DNA:RNA hybrids) to mediate enhancer-promoters interaction (EPI) [136]. R-loops in hESCs are enriched in key cell lineage-specific transcription factors genes, such as the already mentioned NANOG and OCT4 [137].

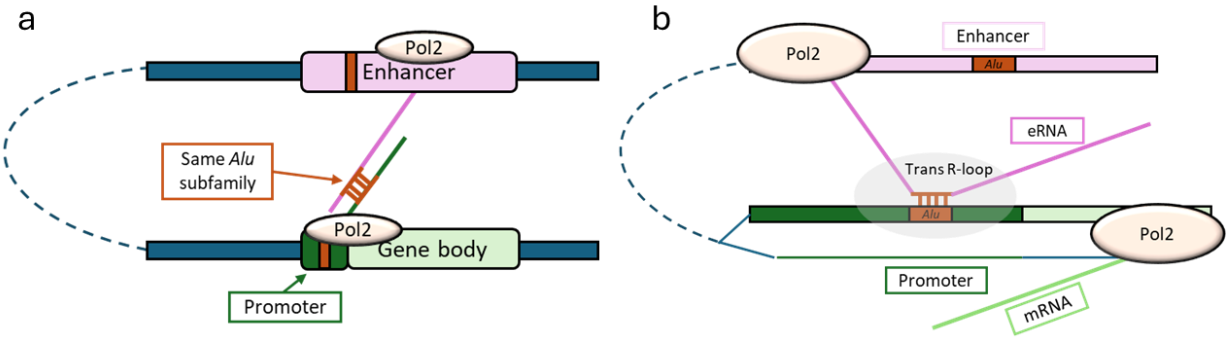

Figure 1. Schematic representation of EPRI (a) and EPI (b). Adapted from [87], [136]

*Alu*s has been reported to be platforms to remodel the human 3D genome structure [5], [138], [139]; *Alu*s flanking NANOG locus, interacting with the transcription factors AhR, lead to the formation of a chromatin loop during the differentiation process, causing a downregulation of NANOG transcription [138]; TFIIIC has also been shown to bind *Alu*s and its occupancy dramatically increase upon serum starvation in several cells line, enabling their H3K18ac and promoting the formation of chromatin loops [5]. TFIIIC-mediated chromatin looping has been reported also in hESCs H9 [139]. In this cell line, about half of TFIIIC-bound regions identified by ChIP-seq are *Alu*s, associated with genes involved in stemness maintenance. Interesting, in

IMR90 cells (which are fibroblasts from human fetal lungs) [Chapter 1, Figure 4], we identified a cluster of genomic regions enriched in *Alu*s-bound by TFIIIC which were associated with genes involved in embryonic development and that was depleted for TFIIIC upon *dl*1500 infection.

The epigenetic profile of stem cell

The main properties that characterize stem cells are the ability to self-renewal and to differentiate into specialized cell types. The capacity to differentiate depends on their plasticity: totipotent cells can differentiate in any cell; pluripotent cells can differentiate in any cell except the placenta and multipotent cells can differentiate in some cellular types [140]. These properties have made stem cells of great interest in regenerative medicine and cancer treatment. Nowadays two multipotent states are generally considered: hESCs and human-induced pluripotent stem cells (iPSCs). Whereas the first is derived from the inner cell mass of the blastocyst, the latter is obtained by somatic cell reprogramming. Given the origin of iPSCs, they pose clinical problems due to their partially differentiated phenotype [141].

The differentiation process requires changes in gene expression profile, which are triggered by alterations in the epigenetic landscape of the cell [142]. Pluripotent cells have a hyperdynamic chromatin state characterized by a lower level of H3K27me3 and a higher level of H3K4me1, but with a lower H3K27ac indicating that many genomic regions could be acting as “poised” enhancers [143], [144]. Despite a lower level of H3K27ac, hESCs chromatin shows high levels of H3K9ac, H3K18ac, and other acetylation marks: the general level of hyperacetylation is reflected in the accessible state of the chromatin of pluripotent cells [145]. The potential to differentiate in different cell types is also reflected in the enrichment in hESCs of bivalent promoters, i.e. the simultaneous presence of H3K27me3 and H3K4me3 [146].

The maintenance of an open state also contributes to the wide distribution of the histone variant H2A.Z. During differentiation, chromatin became less accessible, with a progressive increase in H3K27me3 and H3K4me1. This phenomenon is coupled with an increment of p300/CBP HAT activity during myogenic cell differentiation [147] that lead to the enrichment of H3K27ac and H3K18ac but for more cell type specific loci.

Canonical acetyltransferase: p300/CBP

The acetylation of lysine on histone tails is one of the main post-translation modifications (PTM) to modulate the chromatin state and is generally associated with transcription activation. The canonical lysine acetyltransferases are well characterized and grouped into three main families: the MYST family, the GNAT family and the p300/CBP family. These families share a common acetyl-CoA binding site, whereas diverge in the domain that defines the substrate specificity. GNAT and p300/CBP members may have a bromodomain, whereas MYST members may contain a chromodomain.

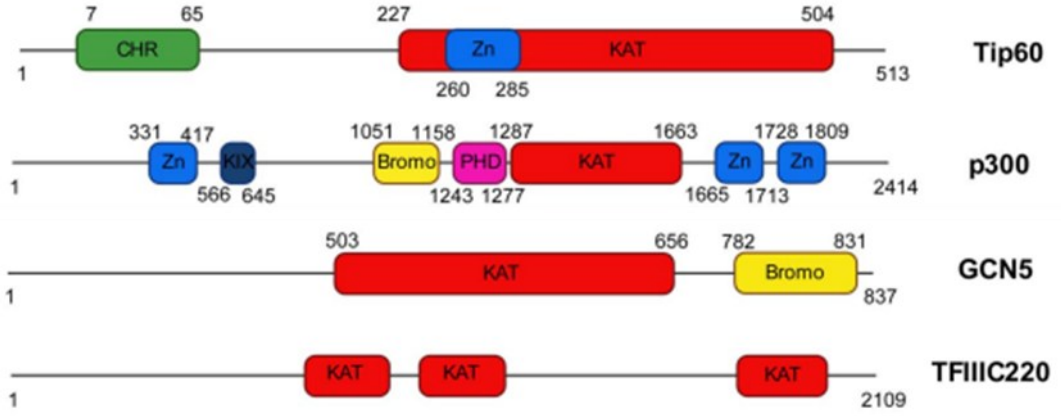

Figure 2. Schematic representation of different KAT families. In the figure are represented examples of the three main classes of KAT and a non-canonical KAT. Tip60 is an example of MYST family. p300 is an example of p300/CBP family. GCN5 is a typical KAT of GNAT family. TFIIIC220 is an example of a non-canonical KAT, as we will discuss in the next paragraph. Adapted from [148]

P300 and CBP are homologous proteins with high sequence similarity and an overlapping structure, commonly referred to as p300/CBP complex. P300 has a molecular weight of ~ 300 kDa and is composed of several domains. The acetyltransferase core is composed of the HAT domain, the bromodomain, the PHD domain and the RING domain, which interact with the auto-inhibitory loop (AIL) inside the HAT domain and has a negative action on the HAT activity. The acetylation state of the AIL regulates the HAT activity of p300: following the interaction with a transcription factor, the hypo-acetylated AIL can insert in the HAT domain of a second p300 molecule and undergoing *in-trans* autoacetylation [149] , explaining how p300, in addition for being at H3K27ac sites, is also present at poised enhancer enriched in H3K27me3. Furthermore, the ability of the hypo-acetylated p300 to form liquid droplets close to

H3K27me3 suggests that a pool of p300 with reduced HAT activity is compartmentalized at poised enhancers [150].

P300/CBP is the main HAT for the acetylation of histone H3 for lysine-18 and lysine-27 in mouse embryonic fibroblasts (MEFs) since are responsible for over 90% of the acetylation [44]. Furthermore, the expression of small e1a, following the adenovirus infection of human embryonic fibroblasts, causes a global hypoacetylation of H3K18 and H3K27 [67], [71]. Since p300/CBP complex plays a major role in the modulation of H3K27ac and H3K18ac during the differentiation process, in differentiated cells and its catalytic activity seems to be limited in hESC [147], it is likely that other HATs could be involved in the deposition of H3K27 and H3K18 acetylation in embryonic stem cells.

Non-canonical acetyltransferase: TFIIIC

Besides the three major families of canonical KAT, other proteins have been reported to possess a lysine acetyltransferase activity [148]. These non-canonical KATs do not have the conserved classical domains but show structure similarity. Among the non-canonical KATs, there are some general transcription factors (GTFs) such as the protein hTAFII250, (the largest subunit of TFIID), which showed HAT activity against H3 and H4 *in vivo* [151]. Also, TFIIB has been found to have autoacetylation activity but did not possess HAT activity *in vitro* [152]. Another GTF, the Pol3 transcription factor TFIIIC have been shown to hold acetyltransferase activity specifically against H3K18 [5], [153] .

TFIIIC is a complex interacting with the Pol3 internal promoter sequences named A and B-box which is conserved among eukaryotes. The yeast TFIIIC complex is composed of the subunits τ131, τ95, and τ55, forming the subcomplex τA, and of τ138, τ91 and τ60 forming the subcomplex τB [154]. The orthologs of human subunits are TFIIIC220, TFIIIC110, TFIIIC90 forming τB and TFIIIC102, TFIIIC63, and TFIIIC35 forming τA. The τA subcomplex recognizes the A-box of the Pol3 promoter, whereas τB subcomplex interacts with the B box. Once bound TFIIIC recruits TFIIIB which in turn recruits the Pol3 [155]. Besides the binding to Pol3 promoter, TFIIIC binds the so-called “extra TFIIIC” sites (ETCs) without the involvement of the transcription machinery [156] and participates in the organization of 3D chromatin structure [157].

The subunits TFIIIC220, TFIIIC110, and TFIIIC90 are involved in the HAT activity of the TFIIIC complex. The HAT activity of these subunits has been characterised *in vitro* through the recombinant expression of these proteins [153], [158], [159]. *In vivo* has been shown that, in a condition of serum starvation, there was an increase of co-occupancy of TFIIIC and H3K18ac on *Alu*s close to Pol2 promoters, but lacking p300 occupancy [5]. The knockdown of TFIIIC220 in different cell lines caused a decrease in the global H3K18ac [5], [153]. These observations suggest that TFIIIC and its newly identified HAT activity could have a more prominent role than previously thought.

Aim of the research

The application of stem cells and induced pluripotent Stem Cells (iPSC) in regenerative medicine and oncology is an actual topic but presents some caveats due to a not full understanding of the epigenetic regulation of "stemness" and their differentiation process. During myogenic cell differentiation there is a significant increase in the HAT activity of p300/CBP [147] (the major lysine acetyltransferase responsible for the acetylation of H3K18 and H3K27 [44]). Thus far, studies on the different histone acetyltransferase active in hESC have not been conducted, so is unknown if p300/CBP represent the major player in the H3K18 and H3K27 acetylation, or if there are other HATs capable of acetylating the same residues. Recently TFIIIC220 has been identified as a HAT specific against H3K18 [5], [153]. Moreover, its activity seems to increase in cells grown in serum-free media [5]. Interestingly the most common medium used for culturing iPSC and hESCs cells (mTeSR and E8) are serum-free [160]. Given these results It is possible to hypothesize, that p300/CBP activity against acetylation of H3K18 and H3K27 in stem cell might be reduced, and TFIIIC could potentially make up and play a major role in the acetylation of H3K18ac. To investigate this, we generated ChIP-seq data of TFIIIC genome-wide occupancy in human embryonic stem cells (hESC) H9 and compared this to published datasets for genome-wide distribution of H3K18ac, H3K27ac, p300 and TFIIIC in hESCs, IMR90 and hESCs-derived neuroectodermal cells.

Methods

External Data Sourses

ChIP-seq data for p300, H3K18ac, H3K27ac in H1, H9, IMR90 and NEC cells were taken from GEO: GSE16256, GSE17917, GSE43152, GSE24447. Chip-seq data for TFIIIC in H9 was from GSE195499, while TFIIIC ChIP-seq data in IMR90 were from GSE120162.

ChIP-seq data analysis

ChIP-seq data analysis was carried out as previously described [72]. Minor modification were from [161]. The average ChIP-seq signal and heatmap profiles were visualized using the tools plotHeatmap from the deepTools package v3.5.1 [96]. Bed intersection was carried out using the “intersectBed” function of the BEDtools program package v2.29.1 with default parameters of 1-bp overlap and custom R scripts.

Cistrome ToolKit analysis (GIGGLE score)

Peaks were used for calculating GIGGLE score [112] by using “all peaks” and “chromatin and transcription factor regulation” options in toolkit analysis of Cistrome Data Browser [108] (accessed on 29 December 2022).

Gene ontology (GO) analysis

g:profiler [162] (accessed on 29 December 2022) was used with default parameters to detect GO terms enrichments and associations of peaks with genes. Lists of genome coordinates derived from downstream analysis of ChIP-seq data were analysed with g:profiler using default parameters.

Results

H3K18ac and H3K27ac do not overlaps with p300 occupancy in hESCs

To understand the relationship between p300 and the deposition of the epigenetics marks H3K18ac and H3K27ac during hESCs differentiation, we analysed ChIP-seq data of these factors in cell lines with at different differentiation states. We compared H3K18ac and p300 ChIP-seq data derived from H1 hESCs and IMR90 cells and for each cell line we ranked the significantly enriched H3K18ac peaks and plotted the p300 signal for the same regions [Figure 1a]. We determined that only 20% of H3K18ac regions were occupied by p300 in H1 hESCs, whereas IMR90 presented 75% of overlapping [Figure 1c]. The same analysis was performed with ChIP-seq data of H3K27ac and p300 collected in Neuroectodermal cells (NECs; these cells are less differentiated than IMR90) and H9 hESCs, ranking the H3K27ac regions from high to low [Figure 1b]. The overlap between H3K27ac and p300 occupancy was ~25 % in H9 hESCs and ~50% for the partially differentiated NECs [Figure 1d]

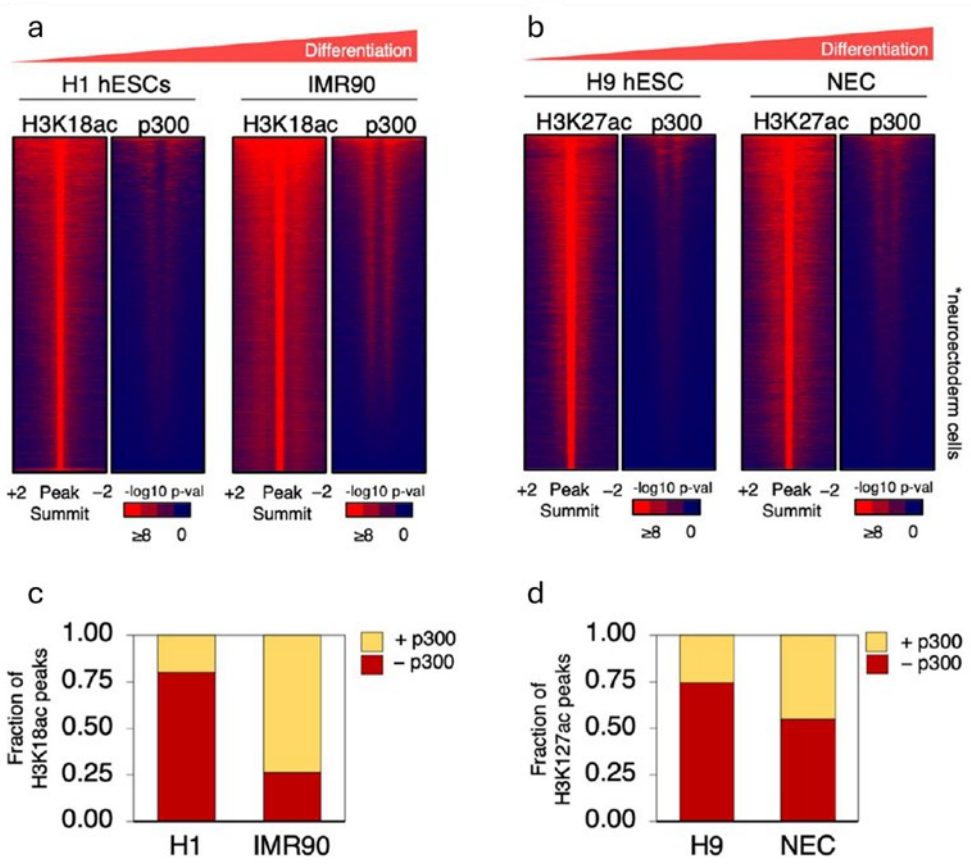

Figure 1. (on previous page) p300 is less associated with H3K18ac and H3K27ac in hESCs than in differentiated cells. (a) Heatmaps of the enrichment of H3K18ac ChIP-seq performed in H1 hESCs and IMR90 cells. Regions are independently ranked from high to low acetylation. The same regions are represented in the heatmaps associated with the ChIP-seq enrichment of p300. Color bar scale with increasing shades of colors stands for higher enrichment (-log10 of the Poisson P-value)**. (b)** Heatmaps of the enrichment of H3K27ac ChIP-seq performed in H9 hESCs and NECs. Regions are independently ranked from high to low acetylation. The same regions are represented in the heatmaps associated with the ChIP-seq enrichment of p300. **(c)** Fraction of H3K18ac peaks occupied (+p300) or not (-p300) for the indicated cell lines represented in (a). **(d)** Fraction of H3K27ac peaks occupied (+p300) or not (-p300) for the indicated cell lines represented in (b).

Taken together this data show that genome-wide H3K18ac or H3K27ac better colocalize with p300 with increasing differentiation. In addition, about the ~75-80% of H3K18 and H3K27 acetylation in hESCs was found without p300 binding, therefore suggesting that other HATs might be involved in the depositions of these marks in hESCs.

TFIIIC220 as a putative H3K18 HAT in hESCs

TFIIIC220, the largest subunit of TFIIIC, shows HAT activity against H3K18 *in vitro* and *in vivo*, which is increase upon serum starvation in T47D cells. Since hESCs are generally growth in serum-free media and given the observed discrepancy between p300 occupancy and H3K18ac and H3K27ac from hESCs to differentiated cells [Figure 1], we decided to perform TFIIIC ChIP-seq in H9 hESCs and compared it with H3K18ac and H3K27ac ChIP-seq data. In H9 hESCs, H3K18ac colocalized with TFIIIC but not p300 in ~1760 regions. In these regions H3K27ac was also less represented than H3K18ac [Figure 2a]. The Gene Ontology analysis revealed that these loci were enriched in genes involved in the maintenance of pluripotency in hESCs (GO: Wikipathways, [Figure 2b, upper panel]) and in neurogenesis (GO: Cellular component, [Fig.2b, bottom panel]). As exemplified by the genomic loci CHD1, HELLS and KLF4, TFIIIC and H3K18ac were lost upon differentiation, nevertheless there was an increase in p300/CBP occupancy and H3K18ac at a marker of differentiated fibroblast, like the COL6A3 gene [Figure 3]. These results supported the hypothesis that TFIIIC could have a role in maintaining the pluripotency of hESCs. Furthermore, the association of TFIIIC with genes involved in the neuronal differentiation, suggested a potential involvement of TFIIIC in the neuronal development. The Cistrome Toolkit analysis [112] was employed to search all the available ChIP-seq on the CistromeDB for other putative transcription factors and chromatin regulators occupying the

TFIIIC-bound regions devoid of p300 in H9 hESCs [Figure 2d, left]. We found that in addition to some components of the Pol 3 machinery, many factors implied in neuronal biology resulted to be significantly enriched. We found two components of ChAHP complex (CHD4 and ADNP), involved in the regulation of 3D chromatin shape and associated to neuronal development [163]; components of complexes PCGF1-PRC1 (RNF2, KDM2B, RYBP and PCGF1) and PRC2.2 (EZH1/2, and JARID), both regulators of neurogenesis and generally of cell differentiation [164]. ChAHP and PRC2 complexes are also known interactors of TFIIIC [5], [165]. The cistrome Toolkit analysis on the regions where H3K18ac colocalized with TFIIIC, showed enrichment for many factors involved in the Pol2 transcription such as POLR2A (the major subunit of Pol2), BRD2, BRD4, NELFA and NELFE [Figure 2d, right]. Taken together these data suggested that TFIIIC might participate in the transcription of genes associated with stem cell pluripotency through the acetylation of H3K18, while at the same time occupying genes required for neuronal development prior to differentiation.

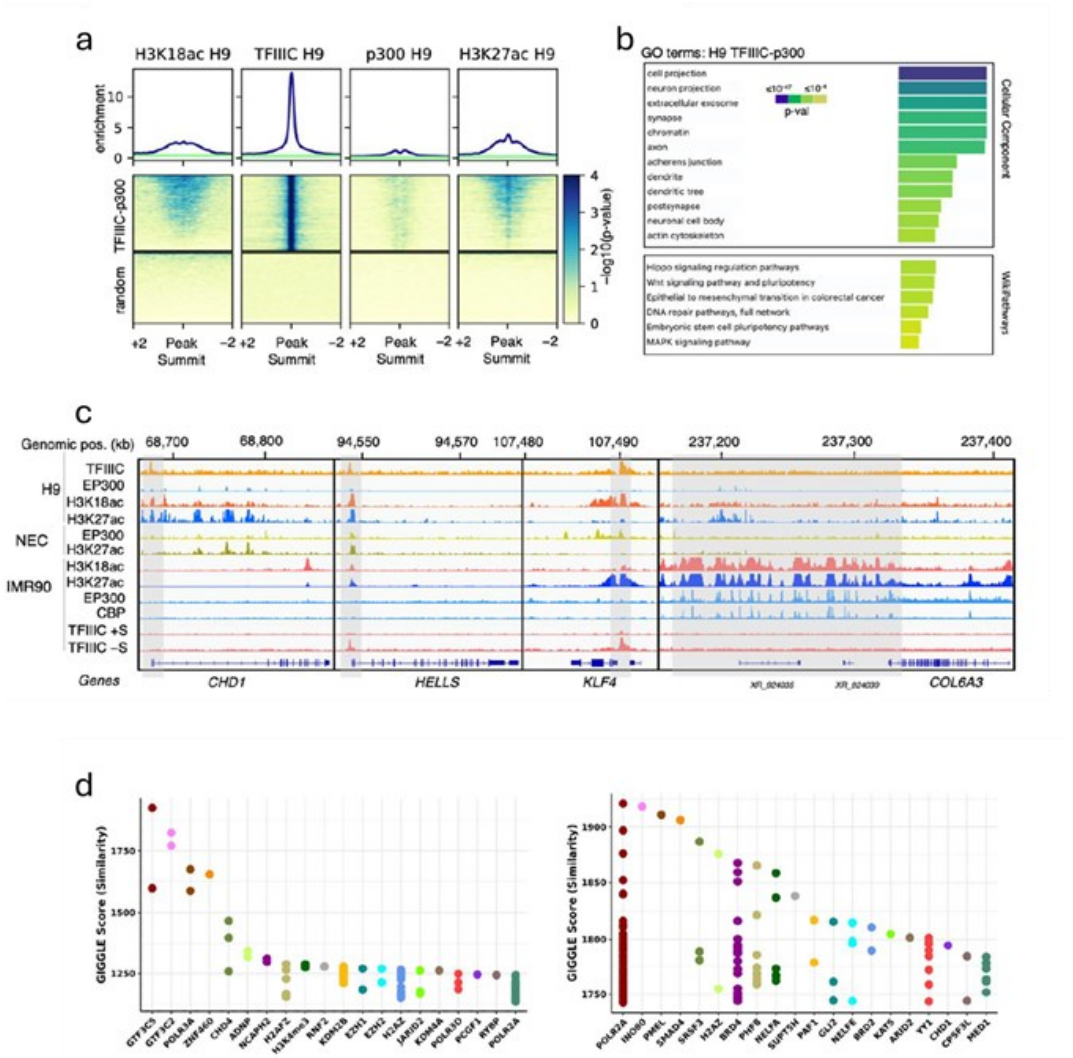

Figure 2. TFIIIC binds close to promoters of stem cells genes and neuronal genes. (a) Average levels and heatmaps for H3K18ac, TFIIIC, p300 and H3K27ac in H9 hESCs. The panels on top of the heatmaps show the enrichment of the indicated factors at TFIIIC peaks, as seen in the corresponding heatmaps (medium panels). Note that H3K18ac and H3K27ac loci positive for TFIIIC are devoid of p300. The lower panels are heatmaps showing the enrichment of the indicated factors for random set of peaks. **(b)** g:Profiler analysis[162] of TFIIIC peaks in H9 hESCs, showing GO terms enriched for cellular component and Wikipathways. The p-value for enrichment is shown as a colour code. **(c)** Genome browser examples of H3K18ac and H3K27ac peaks in H9 hESCs, NECs and IMR90 fibroblasts: stem cell-associated genes such as *CHD1*, *HELLS*, *KLF4*, and the fibroblast- expressed marker *COL6A3* gene are shown. These stem cell loci are devoid of p300 but associated with TFIIIC in H9 hESCs (the grey bars indicate TFIIIC peaks within the promoter regions of these genes). In contrast, *COL6A3* is associated withp300/CBP in IMR90 cells. At the promoters of *KLF4* and *HELLS*, the TFIIIC peaks is also visible in serum-starved IMR90 cells (-S). **(d)** GIGGLE score analysis[112] for all TFIIIC peaks (left) and TFIIIC peaks overlapping H3K18ac peaks (right) in hESCs**.**

Discussion

The ChIP-seq datasets analysed in this study suggest the p300/CBP in likely not be the only HAT involved in H3K18ac and H3K27ac deposition. An increase in the co-localization of these factors was instead observed in NECs and IMR90. The observation that in hESCs TFIIIC co-mapped with many H3K18ac sites where p300 was absent suggested that TFIIIC could be in part responsible for the acetylation, in virtue of its reported HAT activity [5], [153].

TFIIIC, in addition to be required for Pol3 transcription, it is also involved in chromatin organization [5]. This last property, in addition to the HAT activity, supports the hypothesis that TFIIIC is involved in the transcriptional modulation of those genes that are involved in neuronal differentiation. Interestingly, we found that in hESCs TFIIIC occupied-regions associated with neurogenesis lacking H3K18ac, whereas this acetylation was present in regions associated with genes involved in the maintenance of pluripotency in hESCs. Notably, CRISPR-Cas9 screening [133], [134], identified TFIIIC220 as an important gene for the maintenance of pluripotency. The results obtained in our work therefore suggest that TFIIIC might be involved in the modulation of the transcription of those genes involved in neurodifferentiation and pluripotency, through its HAT activity in addition to its role in 3D genome organization.

The results presented in this study are correlative and do not exclude the presence of other enzymes with HAT activity responsible for the acetylation of H3K18. Further studies are required to determine the role of TFIIIC220 HAT activity in hESCs. *In vitro* assays identified mutations of the catalytic domain of TFIIIC220 that reduced the acetyltransferase activity [151]. The small molecule embelin has been identified as a TFIIIC220 inhibitor [151] but lacks specificity since is a recognized PCAF inhibitor [164]. The establishment of an hESC line expressing a mutant TFIIIC220 lacking HAT activity could unveil the role of TFIIIC220 in the maintenance of pluripotency and in neuronal differentiation.

In conclusion, our study suggests that p300/CBP are not the only HATs involved in deposition of H3K18ac and H3K27ac in hESCs. The identification of new players in hESCs epigenetic modulation could open new therapeutic avenues in cancer and developmental disorders.

Bibliography

[1] S. Lander *et al.*, "International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome," *Nature*, vol. 409, pp. 860–921, 2001, doi: https://doi.org/10.1038/35057062.

[2] B. Xia *et al.*, "On the genetic basis of tail-loss evolution in humans and apes," *Nature*, vol. 626, no. 8001, pp. 1042–1048, Feb. 2024, doi: 10.1038/s41586-024-07095-8.

[3] S. J. Klein and R. J. O'Neill, "Transposable elements: genome innovation, chromosome diversity, and centromere conflict," *Chromosome Research*, vol. 26, no. 1–2, pp. 5–23, Mar. 2018, doi: 10.1007/s10577-017-9569-5.

[4] P. D. Mariner *et al.*, "Human Alu RNA Is a Modular Transacting Repressor of mRNA Transcription during Heat Shock," *Mol Cell*, vol. 29, no. 4, pp. 499–509, Feb. 2008, doi: 10.1016/j.molcel.2007.12.013.

[5] R. Ferrari *et al.*, "TFIIIC Binding to Alu Elements Controls Gene Expression via Chromatin Looping and Histone Acetylation," *Mol Cell*, vol. 77, no. 3, pp. 475-487.e11, Feb. 2020, doi: 10.1016/j.molcel.2019.10.020.

[6] M. Han, M. H. Perkins, L. S. Novaes, T. Xu, and H. Chang, "Advances in transposable elements: from mechanisms to applications in mammalian genomics," *Front Genet*, vol. 14, Nov. 2023, doi: 10.3389/fgene.2023.1290146.

[7] J. O. Kriegs, G. Churakov, J. Jurka, J. Brosius, and J. Schmitz, "Evolutionary history of 7SL RNA-derived SINEs in Supraprimates," *Trends in Genetics*, vol. 23, no. 4, pp. 158–161, Apr. 2007, doi: doi:10.1016.

[8] Y. Quentin, "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*, vol. 20, no. 3, pp. 487–493.

[9] E. A. Bennett *et al.*, "Active Alu retrotransposons in the human genome," *Genome Res*, vol. 18, no. 12, pp. 1875–1883, Dec. 2008, doi: 10.1101/gr.081737.108.

[10] R. Cordaux, D. J. Hedges, S. W. Herke, and M. A. Batzer, “Estimating the retrotransposition rate of human Alu elements,” *Gene*, vol. 373, no. 1–2, pp. 134–137, May 2006, doi: 10.1016/j.gene.2006.01.019.

[11] A. Conti, D. Carnevali, V. Bollati, S. Fustinoni, M. Pellegrini, and G. Dieci, “Identification of RNA polymerase III-transcribed Alu loci by computational screening of RNA-Seq data,” *Nucleic Acids Res*, vol. 43, no. 2, pp. 817–835, Jan. 2015, doi: 10.1093/nar/gku1361.

[12] E. N. Kroutter, V. P. Belancio, B. J. Wagstaff, and A. M. Roy-Engel, “The RNA Polymerase Dictates ORF1 Requirement and Timing of LINE and SINE Retrotransposition,” *PLoS Genet*, vol. 5, no. 4, p. e1000458, Apr. 2009, doi: 10.1371/journal.pgen.1000458.

[13] A. Orioli, C. Pascali, A. Pagano, M. Teichmann, and G. Dieci, “RNA polymerase III transcription control elements: Themes and variations,” *Gene*, vol. 493, no. 2, pp. 185–194, Feb. 2012, doi: 10.1016/j.gene.2011.06.015.

[14] R. J. White, “Transcription by RNA polymerase III: more complex than we thought,” *Nat Rev Genet*, vol. 12, no. 7, pp. 459–463, Jul. 2011, doi: 10.1038/nrg3001.

[15] A. Gjidoda and R. W. Henry, “RNA polymerase III repression by the retinoblastoma tumor suppressor protein,” *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, vol. 1829, no. 3–4, pp. 385–392, Mar. 2013, doi: 10.1016/j.bbagrm.2012.09.011.

[16] N. S. Kenneth, B. A. Ramsbottom, N. Gomez-Roman, L. Marshall, P. A. Cole, and R. J. White, “TRRAP and GCN5 are used by c-Myc to activate RNA polymerase III transcription,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 38, pp. 14917–14922, Sep. 2007, doi: 10.1073/pnas.0702909104.

[17] I. Chesnokov and C. W. Schmid, “Flanking Sequences of an Alu Source Stimulate Transcription In Vitro by Interacting with Sequence-Specific Transcription Factors,” 1996.

[18] P. Agarwal *et al.*, “Growth signals employ CGGBP1 to suppress transcription of Alu-SINEs,” *Cell Cycle*, vol. 15, no. 12, pp. 1558–1571, Jun. 2016, doi: 10.4161/15384101.2014.967094.

[19] X. O. Zhang, T. R. Gingeras, and Z. Weng, “Genome-wide analysis of polymerase III–transcribed Alu elements suggests cell-type–specific enhancer function,” *Genome Res*, vol. 29, no. 9, pp. 1402–1414, 2019, doi: 10.1101/gr.249789.119.

[20] T.-H. Li and C. W. Schmid, “Differential stress induction of individual Alu loci: implications for transcription and retrotransposition,” *Gene*, vol. 276, no. 1–2, pp. 135–141, Oct. 2001, doi: 10.1016/S0378-1119(01)00637-0.

[21] J. van Arensbergen *et al.*, “Genome-wide mapping of autonomous promoter activity in human cells,” *Nat Biotechnol*, vol. 35, no. 2, pp. 145–153, Feb. 2017, doi: 10.1038/nbt.3754.

[22] K. E. Paulson and C. W. Schmid, “Transcriptional inactivity of Alu repeats in HeLa cells,” *Nucleic Acids Res*, vol. 14, no. 15, pp. 6145–6158, 1986, doi: 10.1093/nar/14.15.6145.

[23] A. J. Oler *et al.*, “Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors,” *Nat Struct Mol Biol*, vol. 17, no. 5, pp. 620–628, May 2010, doi: 10.1038/nsmb.1801.

[24] A. J. Oler, S. Traina-Dorge, R. S. Derbes, D. Canella, B. R. Cairns, and A. M. Roy-Engel, “Alu expression in human cell lines and their retrotranspositional potential,” *Mob DNA*, vol. 3, no. 1, p. 11, Dec. 2012, doi: 10.1186/1759-8753-3-11.

[25] H. Xie *et al.*, “High-throughput sequence-based epigenomic analysis of Alu repeats in human cerebellum,” *Nucleic Acids Res*, vol. 37, no. 13, pp. 4331–4340, 2009, doi: 10.1093/nar/gkp393.

[26] R. J. Klose and A. P. Bird, “Genomic DNA methylation: the mark and its mediators,” *Trends Biochem Sci*, vol. 31, no. 2, pp. 89–97, Feb. 2006, doi: 10.1016/j.tibs.2005.12.008.

[27] A. D. Ewing *et al.*, “Nanopore Sequencing Enables Comprehensive Transposable Element Epigenomic Profiling,” *Mol Cell*, vol. 80, no. 5, pp. 915-928.e5, Dec. 2020, doi: 10.1016/j.molcel.2020.10.024.

[28] W.-M. Liu and C. W. Schmid, “Proposed roles for DNA methylation in *Alu* transcriptional repression and mutational inactivation,” *Nucleic Acids Res*, vol. 21, no. 6, pp. 1351–1359, 1993, doi: 10.1093/nar/21.6.1351.

[29] W.-M. Liu, R. J. Maraia, C. M. Rubin, and C. W. Schmid, “Alu transcripts: cytoplasmic localisation and regulation by DNA methylation,” *Nucleic Acids Res*, vol. 22, no. 6, pp. 1087–1095, 1994, doi: 10.1093/nar/22.6.1087.

[30] A. Daskalos *et al.*, “Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer,” *Int J Cancer*, vol. 124, no. 1, pp. 81–87, Jan. 2009, doi: 10.1002/ijc.23849.

[31] M. Jordà *et al.*, “The epigenetic landscape of *Alu* repeats delineates the structural and functional genomic architecture of colon cancer cells,” *Genome Res*, vol. 27, no. 1, pp. 118–132, Jan. 2017, doi: 10.1101/gr.207522.116.

[32] D. Varshney, J. Vavrova-Anderson, A. J. Oler, V. H. Cowling, B. R. Cairns, and R. J. White, “SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation,” *Nat Commun*, vol. 6, no. 1, p. 6569, Mar. 2015, doi: 10.1038/ncomms7569.

[33] X. Xiong, H. Chen, Q. Zhang, Y. Liu, and C. Xu, “Uncovering the roles of DNA hemi-methylation in transcriptional regulation using MspJI-assisted hemi-methylation sequencing,” *Nucleic Acids Res*, vol. 52, no. 5, pp. e24–e24, Mar. 2024, doi: 10.1093/nar/gkae023.

[34] J. S. Becker, D. Nicetto, and K. S. Zaret, “H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes,” *Trends in Genetics*, vol. 32, no. 1, pp. 29–41, Jan. 2016, doi: 10.1016/j.tig.2015.11.001.

[35] M. Su, D. Han, J. Boyd-Kirkup, X. Yu, and J.-D. J. Han, “Evolution of Alu Elements toward Enhancers,” *Cell Rep*, vol. 7, no. 2, pp. 376–385, Apr. 2014, doi: 10.1016/j.celrep.2014.03.011.

[36] J. Yan *et al.*, “Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers,” *Cell Res*, vol. 28, no. 2, pp. 204–220, Feb. 2018, doi: 10.1038/cr.2018.1.

[37] A. Local *et al.*, “Identification of H3K4me1-associated proteins at mammalian enhancers,” *Nat Genet*, vol. 50, no. 1, pp. 73–82, Jan. 2018, doi: 10.1038/s41588-017-0015-6.

[38] L. Rinaldi *et al.*, “Dnmt3a and Dnmt3b Associate with Enhancers to Regulate Human Epidermal Stem Cell Homeostasis,” *Cell Stem Cell*, vol. 19, no. 4, pp. 491–501, Oct. 2016, doi: 10.1016/j.stem.2016.06.020.

[39] M. P. Creyghton *et al.*, “Histone H3K27ac separates active from poised enhancers and predicts developmental state,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 50, pp. 21931–21936, Dec. 2010, doi: 10.1073/pnas.1016071107.

[40] J.-E. Lee *et al.*, “Brd4 binds to active enhancers to control cell identity gene induction in adipogenesis and myogenesis,” *Nat Commun*, vol. 8, no. 1, p. 2217, Dec. 2017, doi: 10.1038/s41467-017-02403-5.

[41] T. Zhang, Z. Zhang, Q. Dong, J. Xiong, and B. Zhu, “Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells,” *Genome Biol*, vol. 21, no. 1, p. 45, Dec. 2020, doi: 10.1186/s13059-020-01957-w.

[42] W. Li, D. Notani, and M. G. Rosenfeld, “Enhancers as non-coding RNA transcription units: recent insights and future perspectives,” *Nat Rev Genet*, vol. 17, no. 4, pp. 207–223, Apr. 2016, doi: 10.1038/nrg.2016.4.

[43] R. Ahuja and V. Kumar, “Stimulation of Pol III-dependent 5S rRNA and U6 snRNA gene expression by AP-1 transcription factors,” *FEBS J*, vol. 284, no. 13, pp. 2066–2077, Jul. 2017, doi: 10.1111/febs.14104.

[44] Q. Jin *et al.*, “Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation,” *EMBO J*, vol. 30, no. 2, pp. 249–262, Jan. 2011, doi: 10.1038/emboj.2010.318.

[45] K. Singh, M. Carey, S. Saragosti, and M. Botchan, “Expression of enhanced levels of small RNA polymerase III transcripts encoded by the B2 repeats in simian virus 40-transformed mouse cells,” *Nature*, vol. 314, no. 6011, pp. 553–556, Apr. 1985, doi: 10.1038/314553a0.

[46] C. G. C. Larminie, J. E. Sutcliffe, K. Tosh, A. G. Winter, Z. A. Felton-Edkins, and R. J. White, “Activation of RNA Polymerase III Transcription in Cells Transformed by Simian Virus 40,” *Mol Cell Biol*, vol. 19, no. 7, pp. 4927–4934, Jul. 1999, doi: 10.1128/MCB.19.7.4927.

[47] J. Karijolich, E. Abernathy, and B. A. Glaunsinger, “Infection-Induced Retrotransposon-Derived Noncoding RNAs Enhance Herpesviral Gene Expression via the NF-κB Pathway,”

PLoS Pathog, vol. 11, no. 11, p. e1005260, Nov. 2015, doi: 10.1371/journal.ppat.1005260.

[48] A. M. Schaller, J. Tucker, I. Willis, and B. A. Glaunsinger, "Conserved Herpesvirus Kinase ORF36 Activates B2 Retrotransposons during Murine Gammaherpesvirus Infection," *J Virol*, vol. 94, no. 14, Jul. 2020, doi: 10.1128/JVI.00262-20.

[49] K. L. Jang and D. S. Latchman, "HSV infection induces increased transcription of Alu repeated sequences by RNA polymerase III," *FEBS Lett*, vol. 258, no. 2, pp. 255–258, Dec. 1989, doi: 10.1016/0014-5793(89)81667-9.

[50] B. Panning and J. R. Smiley, "Activation of RNA Polymerase III Transcription of Human Alu Elements by Herpes Simplex Virus," *Virology*, vol. 202, no. 1, pp. 408–417, Jul. 1994, doi: 10.1006/viro.1994.1357.

[51] S. E. Dremel, F. L. Sivrich, J. M. Tucker, B. A. Glaunsinger, and N. A. DeLuca, "Manipulation of RNA polymerase III by Herpes Simplex Virus-1," *Nat Commun*, vol. 13, no. 1, p. 623, Feb. 2022, doi: 10.1038/s41467-022-28144-8.

[52] T.-H. Li, "K562 cells implicate increased chromatin accessibility in Alu transcriptional activation," *Nucleic Acids Res*, vol. 28, no. 16, pp. 3031–3039, Aug. 2000, doi: 10.1093/nar/28.16.3031.

[53] V. R. Russanova, C. T. Driscoll, and B. H. Howard, "Adenovirus Type 2 Preferentially Stimulates Polymerase III Transcription of *Alu* Elements by Relieving Repression: a Potential Role for Chromatin," *Mol Cell Biol*, vol. 15, no. 8, pp. 4282–4290, Aug. 1995, doi: 10.1128/MCB.15.8.4282.

[54] D. Chang, "Monomeric scAlu and nascent dimeric Alu RNAs induced by adenovirus are assembled into SRP9/14-containing RNPs in HeLa cells," *Nucleic Acids Res*, vol. 24, no. 21, pp. 4165–4170, Nov. 1996, doi: 10.1093/nar/24.21.4165.

[55] B. Panning and J. R. Smiley, "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*, vol. 13, no. 6, pp. 3231–3244, Jun. 1993, doi: 10.1128/mcb.13.6.3231-3244.1993.

[56] B. Panning and J. R. Smiley, "Activation of Expression of Multiple Subfamilies of Human Alu Elements by Adenovirus Type 5 and Herpes Simplex Virus Type 1," *J Mol Biol*, vol. 248, no. 3, pp. 513–524, May 1995, doi: 10.1006/jmbi.1995.0239.

[57] W. K. Hoeffler, R. Kovelman, and R. G. Roeder, “Activation of transcription factor IIIC by the adenovirus E1A protein,” *Cell*, vol. 53, no. 6, pp. 907–920, Jun. 1988, doi: 10.1016/S0092-8674(88)90409-6.

[58] S. Yoshinaga, N. Dean, M. Han, and A. J. Berk, “Adenovirus stimulation of transcription by RNA polymerase III: evidence for an E1A-dependent increase in transcription factor IIIC concentration.,” *EMBO J*, vol. 5, no. 2, pp. 343–354, Feb. 1986, doi: 10.1002/j.1460-2075.1986.tb04218.x.

[59] F. Vetrini and P. Ng, “Gene therapy with helper-dependent adenoviral vectors: Current advances and future perspectives,” *Viruses*, vol. 2, no. 9, pp. 1886–1917, Sep. 2010.

[60] S. Kulanayake and S. K. Tikoo, “Adenovirus core proteins: Structure and function,” *Viruses*, vol. 13, no. 3, Mar. 2021, doi: 10.3390/v13030388.

[61] P. Pelka, J. N. G. Ablack, G. J. Fonseca, A. F. Yousef, and J. S. Mymryk, “Intrinsic Structural Disorder in Adenovirus E1A: a Viral Molecular Hub Linking Multiple Diverse Processes,” *J Virol*, vol. 82, no. 15, pp. 7252–7263, Aug. 2008, doi: 10.1128/JVI.00104-08.

[62] N. Avvakumov, R. Wheeler, J. C. D’Halluin, and J. S. Mymryk, “Comparative Sequence Analysis of the Largest E1A Proteins of Human and Simian Adenoviruses,” *J Virol*, vol. 76, no. 16, pp. 7968–7975, Aug. 2002, doi: 10.1128/jvi.76.16.7968-7975.2002.

[63] G. Wang and A. J. Berk, “In Vivo Association of Adenovirus Large E1A Protein with the Human Mediator Complex in Adenovirus-Infected and -Transformed Cells,” *J Virol*, vol. 76, no. 18, pp. 9186–9193, Sep. 2002, doi: 10.1128/jvi.76.18.9186-9193.2002.

[64] A. J. Berk, “Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus,” *Oncogene*, vol. 24, no. 52, pp. 7673–7685, Nov. 2005, doi: 10.1038/sj.onc.1209040.

[65] X. Liu and R. Marmorstein, “Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor,” *Genes Dev*, vol. 21, no. 21, pp. 2711–2716, Nov. 2007, doi: 10.1101/gad.1590607.

[66] S. M. Frisch and J. S. Mymryk, “Adenovirus-5 E1A: Paradox and paradigm,” *Nat Rev Mol Cell Biol*, vol. 3, no. 6, pp. 441–452, 2002, doi: 10.1038/nrm827.

[67] R. Ferrari *et al.*, “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*, vol. 16, no. 5, pp. 663–676, Nov. 2014, doi: 10.1016/j.chom.2014.10.004.

[68] R. Ferrari, M. Pellegrini, G. A. Horwitz, W. Xie, A. J. Berk, and S. K. Kurdistani, "Epigenetic reprogramming by adenovirus e1a," *Science (1979)*, vol. 321, no. 5892, pp. 1086–1088, Aug. 2008, doi: 10.1126/science.1155546.

[69] J. C. Ferreon, M. A. Martinez-Yamout, H. J. Dyson, and P. E. Wright, "Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein," *Proc Natl Acad Sci U S A*, vol. 106, no. 32, pp. 13260–13265, 2009, [Online]. Available: www.pnas.orgcgidoi10.1073pnas.0906770106

[70] R. H. Goodman and S. Smolik, "CBP/p300 in cell growth, transformation, and development," *Genes Dev.*, vol. 14, no. 13, pp. 1553–1577, 2000, [Online]. Available: www.genesdev.org

[71] G. A. Horwitz, K. Zhang, M. A. McBrian, M. Grunstein, S. K. Kurdistani, and A. J. Berk, "Adenovirus Small e1a Alters Global Patterns of Histone Modification," *Science (1979)*, vol. 321, no. 5892, pp. 1084–1085, Aug. 2008, doi: 10.1126/science.1155544.

[72] R. Ferrari *et al.*, "Reorganization of the host epigenome by a viral oncogene," *Genome Res*, vol. 22, no. 7, pp. 1212–1221, Jul. 2012, doi: 10.1101/gr.132308.111.

[73] M. Fuchs *et al.*, "The p400 Complex Is an Essential E1A Transformation Target," *Cell*, vol. 106, pp. 297–307, 2001.

[74] K. Chen *et al.*, "Structure of the human TIP60 complex," *Nat Commun*, vol. 15, no. 1, p. 7092, Aug. 2024, doi: 10.1038/s41467-024-51259-z.

[75] C. Li *et al.*, "Structure of human TIP60-C histone exchange and acetyltransferase complex," *Nature*, Sep. 2024.

[76] K. W. Jeong, K. Kim, A. J. Situ, T. S. Ulmer, W. An, and M. R. Stallcup, "Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation," *Nat Struct Mol Biol*, vol. 18, no. 12, pp. 1358–1365, Dec. 2011, doi: 10.1038/nsmb.2153.

[77] S. Li, T. Wei, and A. R. Panchenko, "Histone variant H2A.Z modulates nucleosome dynamics to promote DNA accessibility," *Nat Commun*, vol. 14, no. 1, Dec. 2023, doi: 10.1038/s41467-023-36465-5.

[78] M. Altaf *et al.*, "NuA4-dependent acetylation of nucleosomal histones H4 and H2A directly stimulates incorporation of H2A.Z by the SWR1 complex," *Journal of Biological*

Chemistry, vol. 285, no. 21, pp. 15966–15977, May 2010, doi: 10.1074/jbc.M110.117069.

[79] J. H. Park, X.-J. Sun, and R. G. Roeder, "The SANT Domain of p400 ATPase Represses Acetyltransferase Activity and Coactivator Function of TIP60 in Basal p21 Gene Expression," *Mol Cell Biol*, vol. 30, no. 11, pp. 2750–2761, Jun. 2010, doi: 10.1128/mcb.00804-09.

[80] S. Ishii *et al.*, "Genome-wide ATAC-see screening identifies TFDP1 as a modulator of global chromatin accessibility," *Nat Genet*, vol. 56, no. 3, pp. 473–482, Mar. 2024, doi: 10.1038/s41588-024-01658-1.

[81] D. Rajagopalan *et al.*, "TIP60 represses activation of endogenous retroviral elements," *Nucleic Acids Res*, vol. 46, no. 18, pp. 9456–9470, Oct. 2018, doi: 10.1093/nar/gky659.

[82] G. V. Helgason, J. O'Prey, and K. M. Ryan, "Oncogene-induced sensitization to chemotherapy-induced death requires induction as well as deregulation of E2F1," *Cancer Res*, vol. 70, no. 10, pp. 4074–4080, May 2010, doi: 10.1158/0008-5472.CAN-09-2876.

[83] K. A. Tworkowski *et al.*, "Adenovirus E1A targets p400 to induce the cellular oncoprotein Myc," *Proc Natl Acad Sci U S A*, vol. 105, no. 16, pp. 6103–6108, 2008, [Online]. Available: www.pnas.org/cgi/content/full/

[84] C. Kim, C. M. Rubin, and C. W. Schmid, "Genome-wide chromatin remodeling modulates the Alu heat shock response," *Gene*, vol. 276, no. 1–2, pp. 127–133, Oct. 2001, doi: 10.1016/S0378-1119(01)00639-4.

[85] W.-M. Liu, W.-M. Chu, P. V. Choudary, and C. W. Schmid, "Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts," *Nucleic Acids Res*, vol. 23, no. 10, pp. 1758–1765, May 1995, doi: 10.1093/nar/23.10.1758.

[86] F. Di Ruocco *et al.*, "Alu RNA accumulation induces epithelial-to-mesenchymal transition by modulating miR-566 and is associated with cancer progression," *Oncogene*, vol. 37, no. 5, pp. 627–637, Feb. 2018, doi: 10.1038/onc.2017.369.

[87] L. Liang *et al.*, "Complementary Alu sequences mediate enhancer–promoter selectivity," *Nature*, vol. 619, no. 7971, pp. 868–875, Jul. 2023, doi: 10.1038/s41586-023-06323-x.

[88] S. Hardy, M. Kitamura, T. Harris-Stansil, Y. Dai, and M. L. Phipps, “Construction of adenovirus vectors through Cre-lox recombination,” *J Virol*, vol. 71, no. 3, pp. 1842–1849, Mar. 1997, doi: 10.1128/jvi.71.3.1842-1849.1997.

[89] A. Dobin *et al.*, “STAR: ultrafast universal RNA-seq aligner,” *Bioinformatics*, vol. 29, no. 1, pp. 15–21, Jan. 2013, doi: 10.1093/bioinformatics/bts635.

[90] Y. Liao, G. K. Smyth, and W. Shi, “featureCounts: an efficient general purpose program for assigning sequence reads to genomic features,” *Bioinformatics*, vol. 30, no. 7, pp. 923–930, Apr. 2014, doi: 10.1093/bioinformatics/btt656.

[91] D. Carnevali and G. Dieci, “Identification of RNA Polymerase III-Transcribed SINEs at Single-Locus Resolution from RNA Sequencing Data,” *Noncoding RNA*, vol. 3, no. 1, p. 15, Mar. 2017, doi: 10.3390/ncrna3010015.

[92] M. I. Love, W. Huber, and S. Anders, “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2,” *Genome Biol*, vol. 15, no. 12, p. 550, Dec. 2014, doi: 10.1186/s13059-014-0550-8.

[93] A. R. Quinlan and I. M. Hall, “BEDTools: a flexible suite of utilities for comparing genomic features,” *Bioinformatics*, vol. 26, no. 6, pp. 841–842, Mar. 2010, doi: 10.1093/bioinformatics/btq033.

[94] S. Weser *et al.*, “Transcription Factor (TF)-like Nuclear Regulator, the 250-kDa Form of Homo sapiens TFIIIB”, Is an Essential Component of Human TFIIIC1 Activity,” *Journal of Biological Chemistry*, vol. 279, no. 26, pp. 27022–27029, Jun. 2004, doi: 10.1074/jbc.M312790200.

[95] Z. Wang and R. G. Roeder, “Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation.,” *Genes Dev*, vol. 11, no. 10, pp. 1315–1326, May 1997, doi: 10.1101/gad.11.10.1315.

[96] F. Ramírez, F. Dündar, S. Diehl, B. A. Grüning, and T. Manke, “deepTools: a flexible platform for exploring deep-sequencing data,” *Nucleic Acids Res*, vol. 42, no. W1, pp. W187–W191, Jul. 2014, doi: 10.1093/nar/gku365.

[97] T. L. Bailey, J. Johnson, C. E. Grant, and W. S. Noble, “The MEME Suite,” *Nucleic Acids Res*, vol. 43, no. W1, pp. W39–W49, Jul. 2015, doi: 10.1093/nar/gkv416.

[98] O. Fornes *et al.*, “JASPAR 2020: update of the open-access database of transcription factor binding profiles,” *Nucleic Acids Res*, Nov. 2019, doi: 10.1093/nar/gkz1001.

[99] E. Harlow, B. R. Franza, and C. Schley, “Monoclonal antibodies specific for adenovirus early region 1A proteins: extensive heterogeneity in early region 1A products,” *J Virol*, vol. 55, no. 3, pp. 533–546, Sep. 1985, doi: 10.1128/jvi.55.3.533-546.1985.

[100] N. Jones and T. Shenk, “An adenovirus type 5 early gene function regulates expression of other early viral genes.,” *Proceedings of the National Academy of Sciences*, vol. 76, no. 8, pp. 3665–3669, Aug. 1979, doi: 10.1073/pnas.76.8.3665.

[101] U. F. Greber and J. W. Flatt, “Adenovirus Entry: From Infection to Immunity,” *Annu Rev Virol*, vol. 6, no. 1, pp. 177–197, Sep. 2019, doi: 10.1146/annurev-virology-092818-015550.

[102] H. Shin, T. Liu, A. K. Manrai, and X. S. Liu, “CEAS: cis-regulatory element annotation system,” *Bioinformatics*, vol. 25, no. 19, pp. 2605–2606, Oct. 2009, doi: 10.1093/bioinformatics/btp479.

[103] D. Varshney, J. Vavrova-Anderson, A. J. Oler, B. R. Cairns, and R. J. White, “Selective repression of SINE transcription by RNA polymerase III,” *Mob Genet Elements*, vol. 5, no. 6, pp. 86–91, Nov. 2015, doi: 10.1080/2159256X.2015.1096997.

[104] Y. Zhao *et al.*, “NucMap: a database of genome-wide nucleosome positioning map across species,” *Nucleic Acids Res*, vol. 47, no. D1, pp. D163–D169, Jan. 2019, doi: 10.1093/nar/gky980.

[105] J. Wang *et al.*, “ATAC-Seq analysis reveals a widespread decrease of chromatin accessibility in age-related macular degeneration,” *Nat Commun*, vol. 9, no. 1, p. 1364, Apr. 2018, doi: 10.1038/s41467-018-03856-y.

[106] K. Chen *et al.*, “DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing,” *Genome Res*, vol. 23, no. 2, pp. 341–351, Feb. 2013, doi: 10.1101/gr.142067.112.

[107] R. C. McLeay and T. L. Bailey, “Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data,” *BMC Bioinformatics*, vol. 11, no. 1, p. 165, Dec. 2010, doi: 10.1186/1471-2105-11-165.

[108] R. Zheng *et al.*, “Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis,” *Nucleic Acids Res*, vol. 47, no. D1, pp. D729–D735, Jan. 2019, doi: 10.1093/nar/gky1094.

[109] C. B. Gocke and H. Yu, “ZNF198 Stabilizes the LSD1–CoREST–HDAC1 Complex on Chromatin through Its MYM-Type Zinc Fingers,” *PLoS One*, vol. 3, no. 9, p. e3255, Sep. 2008, doi: 10.1371/journal.pone.0003255.

[110] D. J. Owen, E. Aguilar-Martinez, Z. Ji, Y. Li, and A. D. Sharrocks, “ZMYM2 controls human transposable element transcription through distinct co-regulatory complexes,” *Elife*, vol. 12, Nov. 2023, doi: 10.7554/eLife.86669.

[111] F. Zanconato *et al.*, “Transcriptional addiction in cancer cells is mediated by YAP/TAZ through BRD4,” *Nat Med*, vol. 24, no. 10, pp. 1599–1610, Oct. 2018, doi: 10.1038/s41591-018-0158-8.

[112] R. M. Layer, B. S. Pedersen, T. DiSera, G. T. Marth, J. Gertz, and A. R. Quinlan, “GIGGLE: a search engine for large-scale integrated genome analysis,” *Nat Methods*, vol. 15, no. 2, pp. 123–126, Feb. 2018, doi: 10.1038/nmeth.4556.

[113] N. Tasdemir *et al.*, “BRD4 Connects Enhancer Remodeling to Senescence Immune Surveillance,” *Cancer Discov*, vol. 6, no. 6, pp. 612–629, Jun. 2016, doi: 10.1158/2159-8290.CD-16-0217.

[114] The ENCODE Project Consortium, “An integrated encyclopedia of DNA elements in the human genome,” *Nature*, vol. 489, no. 7414, pp. 57–74, Sep. 2012, doi: 10.1038/nature11247.

[115] R. Lister *et al.*, “Human DNA methylomes at base resolution show widespread epigenomic differences,” *Nature*, vol. 462, no. 7271, pp. 315–322, Nov. 2009, doi: 10.1038/nature08514.

[116] B. D. Pope *et al.*, “Topologically associating domains are stable units of replication-timing regulation,” *Nature*, vol. 515, no. 7527, pp. 402–405, Nov. 2014, doi: 10.1038/nature13986.

[117] C. Stein *et al.*, “YAP1 Exerts Its Transcriptional Control via TEAD-Mediated Activation of Enhancers,” *PLoS Genet*, vol. 11, no. 8, p. e1005465, Aug. 2015, doi: 10.1371/journal.pgen.1005465.

[118] H. Qian, M. Zhu, X. Tan, Y. Zhang, X. Liu, and L. Yang, “Super-enhancers and the super-enhancer reader BRD4: tumorigenic factors and therapeutic targets,” *Cell Death Discov*, vol. 9, no. 1, p. 470, Dec. 2023, doi: 10.1038/s41420-023-01775-6.

[119] J. Seo *et al.*, "AP-1 subunits converge promiscuously at enhancers to potentiate transcription," *Genome Res*, vol. 31, no. 4, pp. 538–550, Apr. 2021, doi: 10.1101/gr.267898.120.

[120] D. Savic *et al.*, "Promoter-distal RNA polymerase II binding discriminates active from inactive CCAAT/ enhancer-binding protein beta binding sites," *Genome Res*, vol. 25, no. 12, pp. 1791–1800, Dec. 2015, doi: 10.1101/gr.191593.115.

[121] C. Mertens and R. G. Roeder, "Different Functional Modes of p300 in Activation of RNA Polymerase III Transcription from Chromatin Templates," *Mol Cell Biol*, vol. 28, no. 18, pp. 5764–5776, Sep. 2008, doi: 10.1128/MCB.01262-07.

[122] G. Büchel *et al.*, "Association with Aurora-A Controls N-MYC-Dependent Promoter Escape and Pause Release of RNA Polymerase II during the Cell Cycle," *Cell Rep*, vol. 21, no. 12, pp. 3483–3497, Dec. 2017, doi: 10.1016/j.celrep.2017.11.090.

[123] M. P. Creyghton *et al.*, "Histone H3K27ac separates active from poised enhancers and predicts developmental state," *Proceedings of the National Academy of Sciences*, vol. 107, no. 50, pp. 21931–21936, Dec. 2010, doi: 10.1073/pnas.1016071107.

[124] N. R. Zemke, D. Gou, and A. J. Berk, "Dedifferentiation by adenovirus E1A due to inactivation of Hippo pathway effectors YAP and TAZ," *Genes Dev*, vol. 33, no. 13–14, pp. 828–843, Jul. 2019, doi: 10.1101/gad.324814.119.

[125] A. J. Bannister and T. Kouzarides, "CBP-induced stimulation of c-Fos activity is abrogated by E1A.," *EMBO J*, vol. 14, no. 19, pp. 4758–4762, Oct. 1995, doi: 10.1002/j.1460-2075.1995.tb00157.x.

[126] G. Dieci, A. Conti, A. Pagano, and D. Carnevali, "Identification of RNA polymerase III-transcribed genes in eukaryotic genomes," *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, vol. 1829, no. 3–4, pp. 296–305, Mar. 2013, doi: 10.1016/j.bbagrm.2012.09.010.

[127] F. Di Ruocco *et al.*, "Alu RNA accumulation induces epithelial-to-mesenchymal transition by modulating miR-566 and is associated with cancer progression," *Oncogene*, vol. 37, no. 5, pp. 627–637, Feb. 2018, doi: 10.1038/onc.2017.369.

[128] S. Cantarella *et al.*, "Alu RNA Modulates the Expression of Cell Cycle Genes in Human Fibroblasts," *Int J Mol Sci*, vol. 20, no. 13, p. 3315, Jul. 2019, doi: 10.3390/ijms20133315.

[129] J. E. Sutcliffe, T. R. P. Brown, S. J. Allison, P. H. Scott, and R. J. White, “Retinoblastoma Protein Disrupts Interactions Required for RNA Polymerase III Transcription,” *Mol Cell Biol*, vol. 20, no. 24, pp. 9192–9202, Dec. 2000, doi: 10.1128/MCB.20.24.9192-9202.2000.

[130] C. G. C. Larminie, “Mechanistic analysis of RNA polymerase III regulation by the retinoblastoma protein,” *EMBO J*, vol. 16, no. 8, pp. 2061–2071, Apr. 1997, doi: 10.1093/emboj/16.8.2061.

[131] A. Gjidoda and R. W. Henry, “RNA polymerase III repression by the retinoblastoma tumor suppressor protein,” *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, vol. 1829, no. 3–4, pp. 385–392, Mar. 2013, doi: 10.1016/j.bbagrm.2012.09.011.

[132] C. R. King, A. Zhang, T. M. Tessier, S. F. Gameiro, and J. S. Mymryk, “Hacking the Cell: Network Intrusion and Exploitation by Adenovirus E1A,” *mBio*, vol. 9, no. 3, Jul. 2018, doi: 10.1128/mBio.00390-18.

[133] R. J. Ihry *et al.*, “Genome-Scale CRISPR Screens Identify Human Pluripotency-Specific Genes,” *Cell Rep*, vol. 27, no. 2, pp. 616-630.e6, Apr. 2019, doi: 10.1016/j.celrep.2019.03.043.

[134] O. Shalem *et al.*, “Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells,” *Science (1979)*, vol. 343, no. 6166, pp. 84–87, Jan. 2014, doi: 10.1126/science.1247005.

[135] A. Yilmaz, M. Peretz, A. Aharony, I. Sagi, and N. Benvenisty, “Defining essential genes for human pluripotent stem cells by CRISPR–Cas9 screening in haploid cells,” *Nat Cell Biol*, vol. 20, no. 5, pp. 610–619, May 2018, doi: 10.1038/s41556-018-0088-1.

[136] X. Bai, F. Li, and Z. Zhang, “A hypothetical model of trans-acting R-loops-mediated promoter-enhancer interactions by Alu elements,” *Journal of Genetics and Genomics*, vol. 48, no. 11, pp. 1007–1019, Nov. 2021, doi: 10.1016/j.jgg.2021.07.005.

[137] P. Yan *et al.*, “Genome-wide R-loop Landscapes during Cell Differentiation and Reprogramming,” *Cell Rep*, vol. 32, no. 1, p. 107870, Jul. 2020, doi: 10.1016/j.celrep.2020.107870.

[138] F. J. González-Rico *et al.*, “Alu retrotransposons modulate Nanog expression through dynamic changes in regional chromatin conformation via aryl hydrocarbon receptor,”

Epigenetics Chromatin, vol. 13, no. 1, p. 15, Dec. 2020, doi: 10.1186/s13072-020-00336-w.

[139] L. de Llobet Cucalon *et al.*, "An RNA Polymerase III General Transcription Factor Engages in Cell Type-Specific Chromatin Looping," *Int J Mol Sci*, vol. 23, no. 4, p. 2260, Feb. 2022, doi: 10.3390/ijms23042260.

[140] U. Lakshmipathy and C. Verfaillie, "Stem cell plasticity," *Blood Rev*, vol. 19, no. 1, pp. 29–38, Jan. 2005, doi: 10.1016/j.blre.2004.03.001.

[141] M. C. Puri and A. Nagy, "Concise Review: Embryonic Stem Cells Versus Induced Pluripotent Stem Cells: The Game Is On," *Stem Cells*, vol. 30, no. 1, pp. 10–14, Jan. 2012, doi: 10.1002/stem.788.

[142] E. R. Gibney and C. M. Nolan, "Epigenetics and gene expression," *Heredity (Edinb)*, vol. 105, no. 1, pp. 4–13, Jul. 2010, doi: 10.1038/hdy.2010.54.

[143] E. Meshorer, D. Yellajoshula, E. George, P. J. Scambler, D. T. Brown, and T. Misteli, "Hyperdynamic Plasticity of Chromatin Proteins in Pluripotent Embryonic Stem Cells," *Dev Cell*, vol. 10, no. 1, pp. 105–116, Jan. 2006, doi: 10.1016/j.devcel.2005.10.017.

[144] J. Zhu *et al.*, "Genome-wide Chromatin State Transitions Associated with Developmental and Environmental Cues," *Cell*, vol. 152, no. 3, pp. 642–654, Jan. 2013, doi: 10.1016/j.cell.2012.12.033.

[145] N. V. Bhanu, S. Sidoli, and B. A. Garcia, "Histone modification profiling reveals differential signatures associated with human embryonic stem cell self-renewal and differentiation," *Proteomics*, vol. 16, no. 3, pp. 448–458, Feb. 2016, doi: 10.1002/pmic.201500231.

[146] A. Harikumar and E. Meshorer, "Chromatin remodeling and bivalent histone modifications in embryonic stem cells," *EMBO Rep*, vol. 16, no. 12, pp. 1609–1619, Dec. 2015, doi: 10.15252/embr.201541011.

[147] A. Polesskaya, "CBP/p300 and muscle differentiation: no HAT, no muscle," *EMBO J*, vol. 20, no. 23, pp. 6816–6825, Dec. 2001, doi: 10.1093/emboj/20.23.6816.

[148] S. Srivastava, S. Kumar, R. Bhatt, R. Ramachandran, A. K. Trivedi, and T. K. Kundu, "Lysine Acetyltransferases (KATs) in Disguise: Diseases Implications," *The Journal of Biochemistry*, vol. 173, no. 6, pp. 417–433, May 2023, doi: 10.1093/jb/mvad022.

[149] E. Ortega *et al.*, “Transcription factor dimerization activates the p300 acetyltransferase,” *Nature*, vol. 562, no. 7728, pp. 538–544, Oct. 2018, doi: 10.1038/s41586-018-0621-1.

[150] Y. Zhang *et al.*, “Nuclear condensates of p300 formed though the structured catalytic core can act as a storage pool of p300 with reduced HAT activity,” *Nat Commun*, vol. 12, no. 1, p. 4618, Jul. 2021, doi: 10.1038/s41467-021-24950-8.

[151] C. A. Mizzen *et al.*, “The TAFII250 Subunit of TFIID Has Histone Acetyltransferase Activity,” *Cell*, vol. 87, no. 7, pp. 1261–1270, Dec. 1996, doi: 10.1016/S0092-8674(00)81821-8.

[152] C. H. Choi, M. Hiromura, and A. Usheva, “Transcription factor IIB acetylates itself to regulate transcription,” *Nature*, vol. 424, no. 6951, pp. 965–969, Aug. 2003, doi: 10.1038/nature01899.

[153] M. Basu, R. Bhatt, A. Sharma, R. Boopathi, S. Das, and T. K. Kundu, “The Largest Subunit of Human TFIIIC Complex, TFIIIC220, a Lysine Acetyltransferase Targets Histone H3K18,” *The Journal of Biochemistry*, vol. 175, no. 2, pp. 205–213, Feb. 2024, doi: 10.1093/jb/mvad088.

[154] C. Ducrot, O. Lefebvre, E. Landrieux, J. Guirouilh-Barbat, A. Sentenac, and J. Acker, “Reconstitution of the Yeast RNA Polymerase III Transcription System with All Recombinant Factors,” *Journal of Biological Chemistry*, vol. 281, no. 17, pp. 11685–11692, Apr. 2006, doi: 10.1074/jbc.M600101200.

[155] W. Seifert-Davila, M. Girbig, L. Hauptmann, T. Hoffmann, S. Eustermann, and C. W. Müller, “Structural insights into human TFIIIC promoter recognition,” *Sci Adv*, vol. 9, no. 27, Jul. 2023, doi: 10.1126/sciadv.adh2019.

[156] Z. Moqtaderi *et al.*, “Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells,” *Nat Struct Mol Biol*, vol. 17, no. 5, pp. 635–640, May 2010, doi: 10.1038/nsmb.1794.

[157] C. Pascali and M. Teichmann, “RNA Polymerase III Transcription – Regulated by Chromatin Structure and Regulator of Nuclear Chromatin Organization,” *Subcell Biochem*, vol. 61, pp. 261–287, 2013, doi: 10.1007/978-94-007-4525-4_12.

[158] T. K. Kundu, Z. Wang, and R. G. Roeder, “Human TFIIIC Relieves Chromatin-Mediated Repression of RNA Polymerase III Transcription and Contains an Intrinsic Histone

Acetyltransferase Activity," *Mol Cell Biol*, vol. 19, no. 2, pp. 1605–1615, Feb. 1999, doi: 10.1128/MCB.19.2.1605.

[159] Y.-J. Hsieh, T. K. Kundu, Z. Wang, R. Kovelman, and R. G. Roeder, "The TFIIIC90 Subunit of TFIIIC Interacts with Multiple Components of the RNA Polymerase III Machinery and Contains a Histone-Specific Acetyltransferase Activity," *Mol Cell Biol*, vol. 19, no. 11, pp. 7697–7704, Nov. 1999, doi: 10.1128/MCB.19.11.7697.

[160] T. Ludwig and J. A. Thomson, "Defined, Feeder-Independent Medium for Human Embryonic Stem Cell Culture," *Curr Protoc Stem Cell Biol*, vol. 2, no. 1, Sep. 2007, doi: 10.1002/9780470151808.sc01c02s2.

[161] R. Zaurin *et al.*, "A set of accessible enhancers enables the initial response of breast cancer cells to physiological progestin concentrations," *Nucleic Acids Res*, vol. 49, no. 22, pp. 12716–12731, Dec. 2021, doi: 10.1093/nar/gkab1125.

[162] J. Reimand, M. Kull, H. Peterson, J. Hansen, and J. Vilo, "g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments," *Nucleic Acids Res*, vol. 35, no. suppl_2, pp. W193–W200, Jul. 2007, doi: 10.1093/nar/gkm226.

[163] S. Clémot-Dupont *et al.*, "The ChAHP chromatin remodelling complex regulates neurodevelopmental disorder risk genes to scale the production of neocortical layers," *Biorxiv*, 2024.

[164] M. Corley and K. L. Kroll, "The roles and regulation of Polycomb complexes in neural development," *Cell Tissue Res*, vol. 359, no. 1, pp. 65–85, Jan. 2015, doi: 10.1007/s00441-014-2011-9.

[165] C. Liu *et al.*, "PRC2 regulates RNA polymerase III transcribed non-translated RNA gene transcription through EZH2 and SUZ12 interaction with TFIIIC complex," *Nucleic Acids Res*, vol. 43, no. 13, pp. 6270–6284, Jul. 2015, doi: 10.1093/nar/gkv574.