



HAL
open science

Insertion of Alu elements impacts sequence-mediated nucleosome positioning

Jérémy Barbier, Fabien Sassolas, Cédric Vaillant, Jean-Nicolas Volff, Frédéric G. Brunet, Benjamin Audit

► To cite this version:

Jérémy Barbier, Fabien Sassolas, Cédric Vaillant, Jean-Nicolas Volff, Frédéric G. Brunet, et al.. Insertion of Alu elements impacts sequence-mediated nucleosome positioning. 23ième Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM), Jul 2022, Rennes, France. pp.130-137. hal-03764544

HAL Id: hal-03764544

<https://hal.science/hal-03764544v1>

Submitted on 30 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Insertion of *Alu* elements impacts sequence-mediated nucleosome positioning

Jérémy BARBIER^{1,2}, Fabien SASSOLAS^{1,2}, Cédric VAILLANT¹, Jean-Nicolas VOLFF², Frédéric G. BRUNET² and Benjamin AUDIT¹

¹ ENS de Lyon, CNRS, Laboratoire de Physique, F-69342 Lyon, France

² Institut de Génétique Fonctionnelle de Lyon, Ecole Normale Supérieure de Lyon, CNRS UMR 5242, Univ Claude Bernard Lyon 1, F-69364 Lyon, France

Corresponding author: benjamin.audit@ens-lyon.fr

Abstract *A physical model for the formation of nucleosomes revealed that the DNA sequence codes for regions inhibiting the formation of nucleosomes. More than 1.6 million nucleosome inhibiting energy barriers (NIEBs) were delineated in the human genome and corroborated as in vivo nucleosome depleted regions (NDRs) in experimental nucleosome occupancy profiles. NIEBs are bordered on each side by compact arrays of 2-3 nucleosomes; one third of the human is covered by this intrinsic nucleosomal organization. Mutation profiles and GC content at NIEB loci presented modulations phased with the compact nucleosomal arrays and the comparison between interspecies and intraspecies divergence rates suggested that nucleosome positioning around these intrinsic NDRs was under selection. In all vertebrates analyzed, similar sequence properties were observed indicating that NIEBs with compacted neighboring nucleosomes are in fact ubiquitous. These results suggest the existence of evolutionary mechanisms linked to the intrinsic nucleosomal organization.*

*We propose an evolutionary scenario in human and chimpanzee where the insertion of *Alu* transposable elements give rise to new NIEBs based on the observation that (i) a majority of *Alu* elements are inserted at NIEB borders, (ii) the most recently inserted *Alu* elements are the closest to the NIEBs, (iii) human insertion sites of polymorphic *Alu* elements show no preferential positioning regarding NIEBs and (iv) the species-specific *Alu* elements are mainly positioned at the borders of species-specific NIEBs. We identify transposable elements (TEs) in mouse and pig that are good candidates to follow a similar model of NIEB creation by insertion of TEs. The dynamics of transposition would thus be a general mechanism of chromatin evolution that could explain the evolutionary success of some TE families. This work allows to better apprehend the evolutionary mechanisms responsible for the genome-wide intrinsic nucleosomal organization.*

Keywords DNA sequence-encoded nucleosome ordering; nucleosome depleted regions; *Alu* elements; chromatin evolution.

1 Introduction

In most Eukaryotes, short chromosome segments (~ 150 DNA base pairs) wrap around cores of 8 histone proteins to form nucleosomes as the elementary structural unit of DNA folding within chromatin [1]. The control of nucleosome positioning and occupancy is critical for the regulation of gene transcription, DNA replication, repair and recombination [1, 2]. We developed a simple physical model of nucleosome occupancy based on the computation of the free energy cost of bending a DNA fragment of a given sequence from its natural curvature to the final superhelical structure around the histone core from which, given an average nucleosome coverage (chemical potential), the nucleosome occupancy profile can be derived [3]. Combining the nucleosome occupancy probability profile and the original energy profile, we identify nucleosome inhibiting energy barriers (NIEBs) as the genomic loci where the energy is high enough to induce a nucleosome-depleted region in the nucleosome occupancy profile [3–6]. This methodology was developed based on yeast experimental nucleosome positioning data. When applied to human, the method delineated an impressive ~ 1.6 million NIEBs, demonstrating that NIEBs are an important feature of the human genome [7]. Importantly, we also observe that the model predictions around NIEBs are in very good agreement with *in vitro* nucleosome occupancy data [8]. Not only is a very low nucleosome occupancy observed within the NIEBs but the

compact positioning of 2-3 nucleosomes (nucleosome repeat length (NRL) < 160 bp) at each NIEB border predicted by the physical model is also observed in the experimental data. The physical model indeed captures intrinsic sequence-dependent nucleosome positioning signal as we also observed for *in vitro* data for the yeast genome [3, 5, 6]. Average *in vivo* nucleosome occupancy profiles [8, 9] are also extremely consistent with the model predictions at NIEB loci, showing that intrinsic NIEBs and the 4-6 well positioned nucleosomes at their borders are also meaningful to describe *in vivo* chromatin organization. Hence, the 1kb-sized regions around NIEBs (one NIEB and its 4-6 flanking nucleosomes) covering 38 % of the human genome, correspond to regions of intrinsic nucleosome positioning that are not subject to chromatin remodeling [7]. These regions are thus of particular interest to study the nucleosome-associated genome evolution such as mutational patterns and transposable element (TE) insertion. We observed that these intrinsic NDRs, together with the flanking nucleosomes, are actually encoded in the GC content and are widely distributed along the 22 human autosomes in GC-rich and GC-poor isochores, in early and late replicating regions, in intergenic and genic regions [7]. By comparing rates of inter- and intraspecies divergences, we have brought evidence of complex patterns of positive and negative selection that depend on the local GC content to maintain some optimal difference in GC composition between the AT-rich intrinsic NDRs and the GC-rich well-positioned first flanking nucleosomes. Intrinsic chromatin organization encoded by NIEBs is ubiquitous across vertebrates as (i) a significant density of NIEBs $\sim 0.6-0.7$ kb $^{-1}$ was observed in all vertebrate genomes analyzed (compared to a density < 0.01 kb $^{-1}$ for a random sequence with equal proportions of A, G, C and T) and (ii) oscillation of the GC profiles nearby NIEBs in these genomes is remarkably phased with the one in human [10].

Transposable elements constitute a ubiquitous component of vertebrate genomes; they constitute from $\sim 6\%$ (Tetraodon, a pufferfish with compact genome) to more than half of the genome in zebrafish and opossum, for example [11]. Initially, TEs were described as selfish DNA that simply take advantage of the cell machinery of their host to multiply and populate the genome as innocent bystanders, mostly without any phenotypic effect [12]. TEs are now recognized as major drivers of gene and genome evolution, with roles in biological diversity and speciation [13–15]. Despite the central role that TEs play in shaping the evolution of eukaryotes, the coupling between the nucleosomal array and the dynamics of TEs has not been fully evaluated. In human, it was indeed proposed that there were no preferential nucleosome positioning over most of the genome [16], preventing the analysis of TE integration sites in relation to nucleosomal positioning. However, using *in vitro* reconstruction of chromatin template with well positioned nucleosomes, it was shown that there exists a dependence on chromatin structure for retroviral integration site selectivity [17, 18]. The observation that in metazoans, the 3D genome folding as measured using chromatin conformation capture experiments correlates with the association of repetitive elements of the same family such as *Alu* in human [19], further highlights the importance to fully characterise the relationship between TEs and chromatin organization. Interestingly, we found that many (52%) *Alu* elements (high copy number non-autonomous TEs found only in primates) were inserted flanking a NIEB [7, 10]. Note that since the nucleosome occupancy model used for NIEB detection was developed on yeast data, the fitting of the model parameters was not influenced by the specific base composition of *Alu* elements. The orientation of the *Alu* elements was strongly dependent on which NIEB side they were inserted, excluding a purely random process. *Alu* sequences are mainly sense at the NIEB 5' end and antisense at the NIEB 3' end, so that the body of the *Alu* element is external to the NIEB. This remarkable positioning results from the matching of the polyA tail of the sense (resp. antisense) *Alu* with the polyA (resp. polyT) located at the edges of some of the predicted NIEBs, leading to an asymmetric distribution of polyA and polyT at *Alu*-associated NIEB borders [7, 10]. Since the terminal polyA at the 3' end of *Alu* plays a critical role in the amplification mechanism [20, 21], this raised the question of the link between NIEBs and *Alu* insertions. The fact that a majority (61%) of NIEBs are free on either side of detectable *Alu* elements initially suggested that NIEBs would pre-exist *Alu* integration [7]. However, *Alu* might be involved in the formation of some but not all NIEBs, or other TEs might be associated to NIEBs, suggesting a convergent mechanism of NIEBs formation by TE insertion. Here we compare the history of *Alu* element insertions along the primate evolutionary tree to NIEB predictions in human and chimpanzee, in order to assess to which extent NIEBs simply constitute a favorable substrate for *Alu* insertions or if these insertions contribute

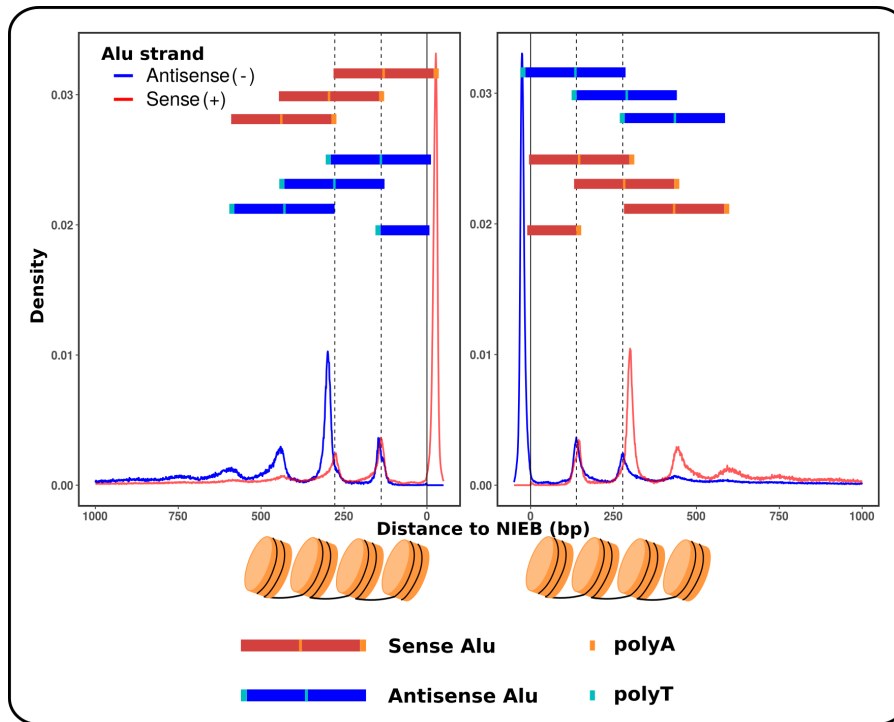


Fig. 1. Distance of the terminal polyA of the 1 078 322 Alu elements on human autosomes to the closest border of nucleosomal barriers (abscissa 0 bp). Alu were classified into 4 categories depending on their orientation (sense Alu in red, antisense Alu in blue) and to which NIEB side they are closest (closest to a NIEB 5' border on the left panel or closest to a NIEB 3' border on the right panel). Each curve is normalized with the total number of Alu of corresponding category (395 076 sense Alu and 142 504 antisense Alu upstream of a NIEB (left panel); 141 548 sense Alu and 398 976 antisense Alu downstream of a NIEB (right panel); 218 Alu were not considered as they fall in a discarded inter-NIEB region containing not sequenced nucleotides (Ns)). Vertical dashed lines mark distances 138 bp and 278 bp from the NIEB borders. The compact arrays of nucleosomes positioned at NIEB borders are drawn under the graph.

to the formation and/or maintenance of NIEBs. In other word, we will address if there is anything else but success of selfishness of the *Alu* amplification to explain the astonishing *Alu* expansion in primates.

2 Results

2.1 Alu elements have specific positioning at NIEBs borders

We first further characterize the positioning of Alu elements relative to NIEBs borders. Alu are dimeric elements, with two GC-rich arms separated by a short polyA sequence and with a longer polyA sequence at their 3' end. To position an Alu element regarding NIEBs, we used the middle of its terminal polyA as a reference position. Figure 1 represents the distributions of these reference positions relative to NIEB borders. Upstream of NIEBs, we observe that sense Alu elements have a strong preferential positioning at NIEBs borders, with most of them being placed with their terminal polyA co-localizing with the NIEB 5' border, as previously described [7, 10]. Interestingly, in this configuration the internal polyA of the elements is placed at the first linker DNA position. Secondary preferential positions are also observed at distance 138 bp and 278 bp from the NIEBs borders, with both the terminal and internal polyAs co-localizing with inter-nucleosomal sequences. The preferential Alu positioning with the polyA in between nucleosomes is also observed for antisense elements, although they represent only a quarter of total Alu elements upstream of NIEBs. Finally, there is a symmetry between the distribution of sense Alu upstream of NIEBs and antisense Alu downstream of NIEBs, and vice-versa.

Alu preferential positioning at NIEBs border could come from several mechanisms:

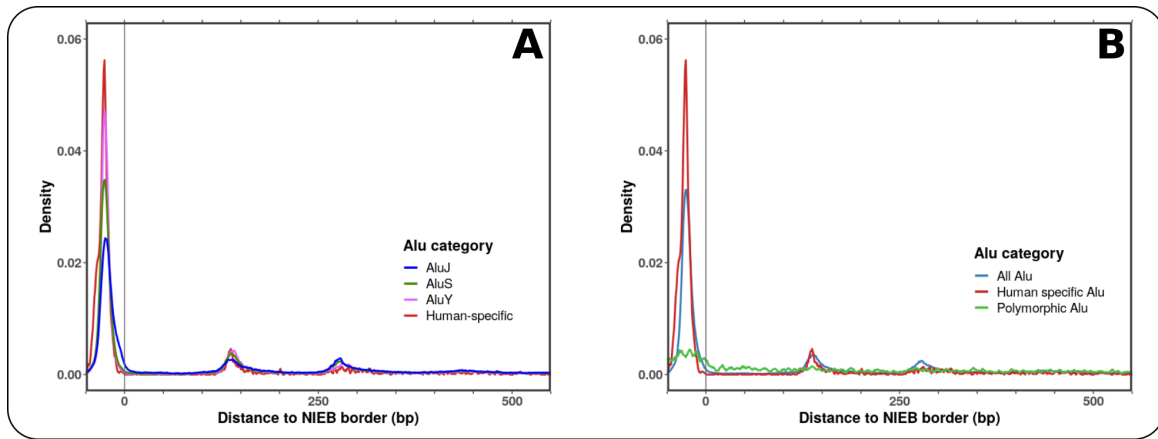


Fig. 2. Distances to the closest nucleosomal barrier border of Alu terminal polyA for different Alu families and of the Alu insertion sites delineated as the location of polymorphic Alu elements from the 1000 Genomes Project. (A) Distance distribution for 95 859 AluJ, 234 121 AluS, 49 188 AluY and 3160 human-specific Alu are represented by the blue, green, pink and red curves, respectively. The analysis is restricted to antisense Alu, less than 2000 bp downstream of a NIEB for which family classification was possible. Alu elements identified as human-specific are included in other categories (mainly in the AluY category). (B) Distance distributions are computed for antisense Alu elements positioned at less than 2000 bp upstream of a NIEB: all 392 807 such Alu in blue, 3160 human-specific Alu in red and 3610 Alu insertion sites of human polymorphic Alu elements in green. Note that x-axis only ranges from -50 bp to 550 bp, because profiles obtained between 550 and 2000 are flat and nearly equal to 0.

- A strong counter-selection of "misplaced" insertions because they would disrupt the chromatin structure.
- Preference for AT-rich, inter-nucleosomal sequences for the insertion of Alu elements because of the insertion mechanism.
- Creation of NIEBs in phase with newly inserted Alu elements, thanks to the inhibitory effect of polyA sequences on nucleosome positioning.

2.2 The latest Alu insertions show strongest constraints on positioning

To explore the hypothesis of a counter-selection of Alu insertions disrupting the chromatin structure, we classified the Alu elements according to their ages (inserting time in the genome). Indeed, we expect the constraints on Alu positioning to be stronger on the oldest categories, simply because the time during which selection has occurred is longer. We used the classification of Alu elements in three main families, with AluJ (26 %) being the oldest elements in the human genome, followed by AluS (59 %), and then AluY (12 %) being the youngest family. We also considered the set of Alu that have been identified as human specific (0.7 %) by Tang et al. [22, 23]. This last category represent the youngest Alu elements of the human genome, as they were inserted after human-chimpanzee divergence. Note that the elements in this category mainly belongs to the AluY family. However, they represent only a small fraction of it (about 6%).

We observe that constraints on Alu positioning are stronger on younger insertions (Figure 2 - A). Indeed, there is a gradient in the proportion of elements positioned at the internal border of NIEBs, with the highest proportion for human-specific Alu, followed by AluY, then AluS and finally AluJ. Thus, the younger an Alu element is, the more its positioning is constrained at NIEBs borders, which is not compatible with the counter-selection hypothesis developed above. Therefore, "misplaced" insertions do not seem to be purified by selection. The next hypothesis is that the insertion mechanism of Alu elements puts them directly where we find them now.

2.3 Human polymorphic Alu insertion sites are not biased toward NIEBs borders

Alu elements are inserted through the binding of their terminal polyA to a pre-existing A-rich sequence in the genome [21]. Yet, an increase in the polyA distribution has been shown at the internal border of NIEBs and at first linker position [7, 10]. This increase could make these regions preferential

insertion platforms for Alu elements, especially as they are also inter-nucleosomal, potentially more accessible to the transposition machinery. To explore this hypothesis, we established the distribution of Alu insertion sites delineated as the location of polymorphic Alu elements extracted from the 1000 Genome Project data [24, 25]. The Alu are present in a human sub-population but are absent from the reference genome. Following our assumption, the distribution of insertion sites should be similar to the one observed for the Alu elements present in the human reference genome.

On Figure 2 - B, we can see that insertion sites of the polymorphic Alu element (in green) show only a very little preference for NIEBs, with only 10.8% of them sitting at the internal border of a NIEB, as compared to the 40% of Alu positioned with their polyA at this locus. This result invalidates the assumption that the insertion mechanism of Alu elements is responsible for their distribution at NIEBs borders.

2.4 Alu insertion creates new nucleosomal barriers

The results presented in sections 2.2 and 2.3 not only invalidated two of our three hypothesis, they also tend to confirm the third one, which stipulates that Alu element insertions would be at the origin of new NIEBs. Indeed, polymorphic elements are not present in human reference genome, and were not taken into account in the human NIEBs prediction. Thus, it seems that NIEBs are found at Alu sites only when the Alu are present in the analyzed sequence. This result is in favor of NIEBs created by the Alu insertions. Moreover, in this model, NIEBs would be formed on the terminal polyA of Alu elements. It is known that this polyA is susceptible to shrinkage when the Alu gets older [26]. Yet, oldest Alu are also less positioned at NIEBs borders. An Alu insertion could then create a NIEB at the terminal polyA of the element. After the insertion, the shrinkage of the polyA could progressively weaken the corresponding NIEB border, which would explain the relaxed constraints on older Alu element positioning at NIEBs borders. To confirm this model of NIEB creation, we compared nucleosome occupancy at Alu sites with and without the inserted Alu sequence using a comparative genomics approach described in section 4.2.

We see that the nucleosome occupancy predicted by our model corroborates our hypothesis of NIEB creation by the insertion of Alu (Figure 3). Indeed, the predicted occupancies in the presence of Alu (in blue) are very different from the ones obtained without the Alu elements (in red). The polyA of the elements (around position 0 bp) systematically form a zone where the occupancy is very low, bordered by two to three well positioned nucleosomes. This nucleosomal organization is characteristic of NIEBs. We also note that the predicted occupancies on Alu elements are particularly high, reflecting the compatibility between these elements and nucleosome positioning. Finally, we observe that the occupancy pattern predicted from our physical model of nucleosome formation is consistent whether or not a NIEB was detected at the insertion site and/or at the border of the Alu element. It is thus clear that the absence of detected NIEB for the Alu of the C and D panels on Figure 3 comes from false negative in NIEBs detection. Hence, a vast majority of human-specific Alu are at the origin of human-specific NIEBs.

3 Discussion

The results presented here demonstrate that the insertion of Alu elements have a strong effect on nucleosomal positioning. Indeed, these elements are at the origin of new NIEBs at the border of which nucleosome are well positioned. This association between Alu elements and NIEBs could have had a role in the spreading of these elements among primates genomes, as they could have been used as spreading factors for NIEBs. Our results also highlight the importance of the terminal polyA of Alu element in NIEB formation. Yet, a lot of SINEs are terminated by polyA, leading to the question of a general interaction between transposable elements and NIEBs. For example, in mouse, B1 elements derives from the same 7SLRNA as Alu elements in primates. The main difference between B1 and Alu is that B1 is a monomer whereas Alu is a dimer. In that way, B1 could be seen as a "half Alu". We found that at the border of NIEBs, B1 are positioned with the same constraints as the ones detailed here for Alu (Figure 1), namely a preference for a positioning with the terminal polyA at the internal border of NIEBs, or at least in an inter-nucleosomal sequence. Thus, B1 elements in mouse could also be involved in NIEBs formation in this genome. In pig, Pre0_SS elements are GC-rich

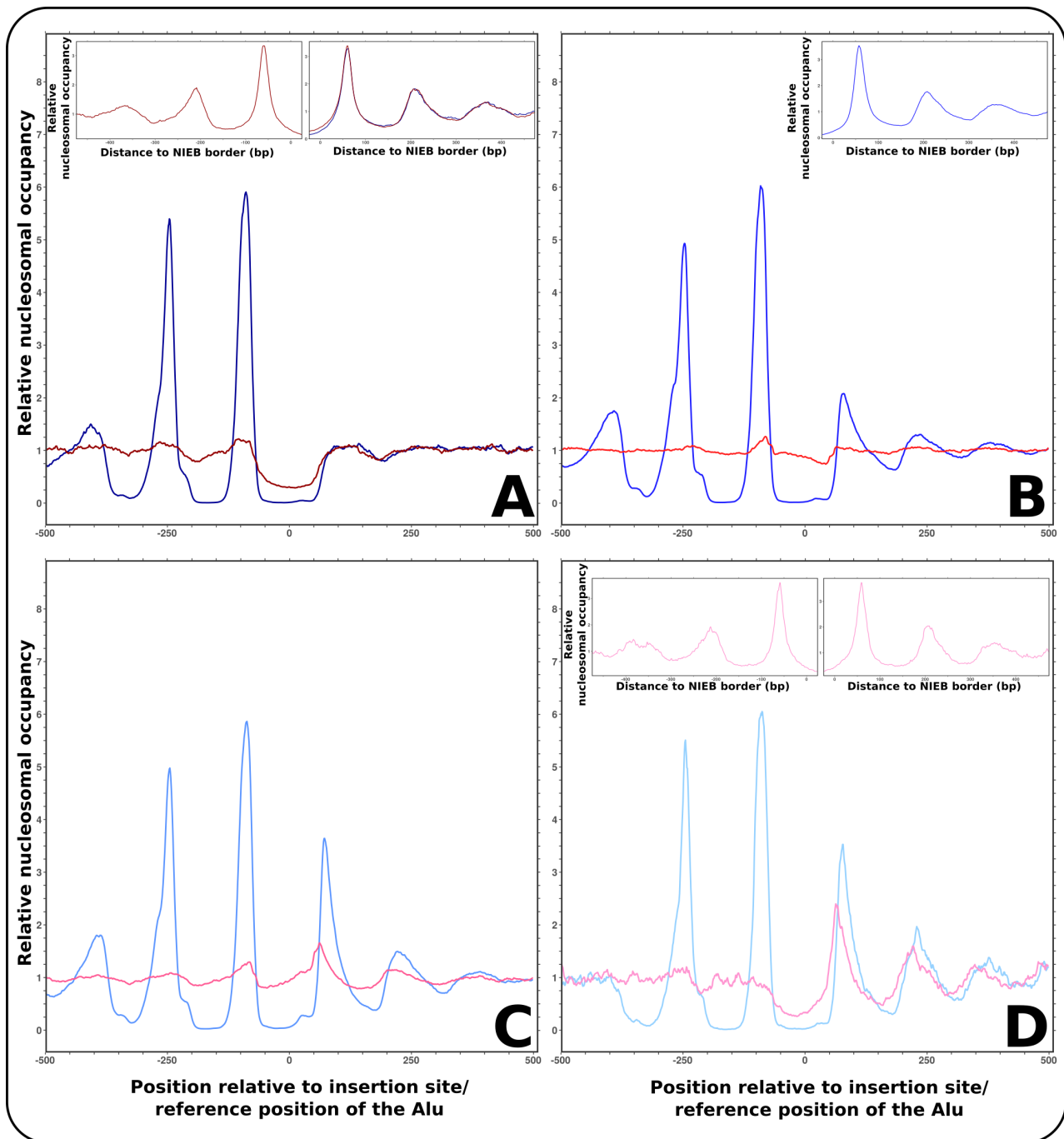


Fig. 3. Mean predicted nucleosomal occupancy at human-specific Alu sites in human and chimpanzee. Red curves represent the profiles obtained in chimpanzee (without Alu element), blue curves the profiles obtained in human (with Alu element). A, B, C and D figures represent respectively cases 1, 2, 3 and 4 described in section 4.2. On each graph, the 0 of the x-axis correspond to reference position for the Alu element in human, and the corresponding insertion site positioned in chimpanzee retrieved through the alignment of the 2x100 bp flanking the human Alu in the chimpanzee genome. Boxes in A, B and D figures represent the same profiles but using as reference position the borders of NIEBs. All these profiles have been normalized with the genomic mean of the signal.

SINEs comprising a terminal polyA. Their size (about 250 bp) and base composition is close to the ones of Alu elements, besides the internal polyA separating the two GC-rich arms of the Alu. Here again, we found that the positioning of these elements at NIEBs borders reproduce the one observed for Alu (and B1) elements. These elements are specific to pigs, and they have a totally different origin as the Alu (they derive from a tRNA-Glu sequence). Thus, it seems that several TE families in several species show similar effect on nucleosomal positioning. This leads to the possibility of a general interaction between NIEBs and TEs, with implication in the evolutionary success of certain TE families. Interestingly, it was observed that regions of preferential DNA replication initiation common to all cell lines were enriched in NIEBs compared to cell line specific replication initiation regions [27, 28], suggesting that TE insertion could contribute to the specification of these constitutive replication initiation regions.

4 Methods

4.1 Reference genomes and annotations

Human (hg38 and hg19) and chimp (panTro5) genomes were downloaded on UCSC database (<http://hgdownload.soe.ucsc.edu/downloads.html>). The annotation of transposable elements used for these two species was the one obtained with RepeatMasker software and provided by the UCSC database. For NIEB prediction, we used the physical model based on sequence-dependent DNA bending properties described in [3–5] to detect 1 745 801, 1 718 916 and 1 733 364 NIEBs respectively on human (hg38 and hg19) and chimp autosomes. *In silico* nucleosome density were obtained with the same method as described in [7] to model high nucleosome coverage *in vivo*. Insertion sites of polymorphic Alu elements were extracted from the set of human polymorphisms detected regarding hg19 genome version in the phase 3 of the 1000 Genomes Project [24, 25]. They were compared to the NIEB prediction obtained for this version of the human genome. NIEBs positions are available at https://perso.ens-lyon.fr/benjamin.audit/Vertebrate_NIEBs/. Predicted nucleosome occupancy profiles can be obtained from the corresponding author.

4.2 Human-specific Alu elements and corresponding insertion sites in chimp

To retrieve ancestral insertion site of human-specific Alu elements, we aligned the flanking regions of each human-specific Alu elements on the chimp genome. 100 bp on each side of each Alu elements were taken and put in fastq format to mimic a paired-end sequencing dataset. These "read pairs" were then aligned on chimp genome using bwa-mem (<http://bio-bwa.sourceforge.net/bwa.shtml>). Only alignments that were properly paired, with the two reads slightly overlapping or being side by side were kept for further analysis. Also, every alignment comprising an indel were discarded, to ensure that the ± 100 bp around the ancestral insertion site were identical as the ± 100 bp around the human Alu element (except for potential SNVs). Human-specific Alu associated with an ancestral insertion site were then separated in 4 groups according to the presence/absence of NIEB at the reference position of Alu element in human and at corresponding insertion site in chimp :

1. A NIEB is present both with (in human) and without (in chimp) Alu element (572 cases, 12.1%)
2. A NIEB is present only associated to the Alu element in human (2229 cases, 47.1%)
3. There is no NIEB neither with nor without the Alu element (1831 cases, 38.7%)
4. A NIEB is present only without the Alu element in chimp (96 cases, 2.0%)

The profiles presented in Figure 3 were obtained by taking the mean occupancy profile of each group in human and chimp, aligned on the reference position.

Acknowledgements

This work was funded by the Agence Nationale de la Recherche (ANR-20-CE12-013), and the ENS de Lyon "Projets Emergents" program. J.B. acknowledges support from the PhD funding program of ENS de Lyon. The authors thank the French CNRS network GDR "Architecture et Dynamique du Noyau et des Génomomes" (ADN&G) for stimulating workshops.

References

- [1] Zlatanova, J. & Van Holde, K. E. (2016). *Molecular biology : structure and dynamics of genomes and proteomes*. Garland Science, Taylor and Francis Group, New-York.

- [2] Arneodo, A., Vaillant, C., Audit, B., Argoul, F., d'Aubenton-Carafa, Y. & Thermes, C. (2011). Multi-scale coding of genomic information: From DNA sequence to genome structure and function. *Phys. Rep.* 498, 45–188.
- [3] Chevereau, G., Arneodo, A. & Vaillant, C. (2011). Influence of the genomic sequence on the primary structure of chromatin. *Front. Life Sci.* 5, 29–68.
- [4] Vaillant, C., Audit, B. & Arneodo, A. (2007). Experiments confirm the influence of genome long-range correlations on nucleosome positioning. *Phys. Rev. Lett.* 99, 218103.
- [5] Milani, P., Chevereau, G., Vaillant, C., Audit, B., Haftek-Terreau, Z., Marilley, M., Bouvet, P., Argoul, F. & Arneodo, A. (2009). Nucleosome positioning by genomic excluding-energy barriers. *Proc Natl Acad Sci U S A* 106, 22257–22262.
- [6] Vaillant, C., Palmeira, L., Chevereau, G., Audit, B., d'Aubenton Carafa, Y., Thermes, C. & Arneodo, A. (2010). A novel strategy of transcription regulation by intragenic nucleosome ordering. *Genome Res* 20, 59–67.
- [7] Drillon, G., Audit, B., Argoul, F. & Arneodo, A. (2016). Evidence of selection for an accessible nucleosomal array in human. *BMC Genomics* 17.
- [8] Valouev, A., Johnson, S. M., Boyd, S. D., Smith, C. L., Fire, A. Z. & Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. *Nature* 474, 516–520.
- [9] Schones, D. E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G. & Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell* 132, 887–898.
- [10] Brunet, F. G., Audit, B., Drillon, G., Argoul, F., Volff, J.-N. & Arneodo, A. (2018). Evidence for DNA Sequence Encoding of an Accessible Nucleosomal Array across Vertebrates. *Biophysical Journal* .
- [11] Chalopin, D., Naville, M., Plard, F., Galiana, D. & Volff, J.-N. (2015). Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.* 7, 567–580.
- [12] Orgel, L. E. & Crick, F. H. (1980). Selfish DNA: the ultimate parasite. *Nature* 284, 604–607.
- [13] McClintock, B. (1956). Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* 21, 197–216.
- [14] Coyne, J. A. & Orr, H. A. (1998). The evolutionary genetics of speciation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 287–305.
- [15] Kraaijeveld, K. (2010). Genome size and species diversification. *Evol. Biol.* 37, 227–233.
- [16] Valouev, A., Johnson, S. M., Boyd, S. D., Smith, C. L., Fire, A. Z. & Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. *Nature* 474, 516–520.
- [17] Pryciak, P. M. & Varmus, H. E. (1992). Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* 69, 769–780.
- [18] Naughtin, M., Haftek-Terreau, Z., Xavier, J., Meyer, S., Silvain, M., Jaszczyszyn, Y., Levy, N., Miele, V., Benleulmi, M. S. *et al.* (2015). DNA physical properties and nucleosome positions are major determinants of HIV-1 integrase selectivity. *PLoS One* 10, e0129427.
- [19] Cournac, A., Koszul, R. & Mozziconacci, J. (2016). The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. *Nucleic Acids Res* 44, 245–255.
- [20] Mighell, A. J., Markham, A. F. & Robinson, P. A. (1997). Alu sequences. *FEBS Lett.* 417, 1–5.
- [21] Deininger, P. (2011). Alu elements: know the SINES. *Genome biology* 12, 236.
- [22] Tang, W., Mun, S., Joshi, A., Han, K. & Liang, P. (2018). Mobile elements contribute to the uniqueness of human genome with 15,000 human-specific insertions and 14 Mbp sequence increase. *DNA Res.* 25, 521–533.
- [23] Tang, W. & Liang, P. (2019). Comparative Genomics Analysis Reveals High Levels of Differential Retrotransposition among Primates from the Hominaidae and the Cercopithecidae Families. *Genome Biol Evol* 11, 3309–3325.
- [24] Durbin, R. M., Altshuler, D., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Collins, F. S., De La Vega, F. M. *et al.* (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073. Number: 7319 Publisher: Nature Publishing Group.
- [25] Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P. *et al.* (2015). A global reference for human genetic variation. *Nature* 526, 68–74. Number: 7571 Publisher: Nature Publishing Group.
- [26] Roy-Engel, A. M., Salem, A.-H., Oyeniran, O. O., Deininger, L., Hedges, D. J., Kilroy, G. E., Batzer, M. A. & Deininger, P. L. (2002). Active Alu element "A-tails": size does matter. *Genome Res* 12, 1333–1344.
- [27] Drillon, G., Audit, B., Argoul, F. & Arneodo, A. (2015). Ubiquitous human 'master' origins of replication are encoded in the DNA sequence via a local enrichment in nucleosome excluding energy barriers. *Journal of Physics: Condensed Matter* 27, 064102.
- [28] Arneodo, A., Drillon, G., Argoul, F. & Audit, B. (2018). The role of nucleosome positioning in genome function and evolution. In *Nuclear Architecture and Dynamics, Vol. 2* (C. Lavelle & J.-M. Victor, eds.), pp. 41–79. Elsevier.