BABEŞ-BOLYAI UNIVERSITY
FACULTY OF BIOLOGY AND GEOLOGY
DOCTORAL SCHOOL OF INTEGRATIVE BIOLOGY

Structural determinants for mobilization of LAVA and SVA VNTR-composites

Doctoral Thesis

Summary

Scientific Supervisor:
Prof. Dr. Nicolaie DRAGOȘ

PhD Candidate:
Maria-Cornelia-Teodora OCHIȘ (CONSTANTINESCU)

Cluj-Napoca
2019

CONTENTS

List of abbreviations

bp	Base pair
cDNA	Complementary DNA
CMV	Cytomegalovirus
C_t	Threshold cycle in qPCR
CT	Cytosine-Thymine rich region
EN	Endonuclease domain
env	Retroviral envelope gene
G418	Geniticin sulfate
HERV	Human Endogenous Retrovirus
Kb	Kilobasepairs
L_ES	LAVA_E-SVA chimera
L_FS	LAVA_F-SVA chimera
mneoI	Retrotransposition reporter cassette
mRNA	Messager RNA
ORF	Open Reading Frame
PGK	Phosphoglycerate Kinase
piwi- RNAs	Suppressors of transposon mobilization
Pol II /pol III	Polymerase II and III
PTBP	Polypyrimidine Tract-Binding Protein
RNP	Ribonucleoprotein
RT	Reverse Transcriptase
SINE-R	SINE of retroviral origin (HERV-K)
SL_E	SVA-LAVA_E chimera
SL_F	SVA-LAVA_F chimera
SVA2	Element in *Macaca mulatta* genome
TPRT	Target Primed Reversed Transcription
TSD	Target Site Duplication
UTR	Untranslated region
VNTR	Variable Number Tandem Repeat
ΔCMV	Without CMV promoter

Keywords

Nomascus leucogenys, gibbon, SVA, LAVA, VNTR, retrotransposition, CT-hexameric repeats.

1. INTRODUCTION

1.1. General introduction

Mobile elements were first discovered by Barbara McClintock in the maize genome (McClintock 1950). In the beginning, they were called "junk DNA" (Ohno 1972), "selfish genes" (Dawkins 1976) and "genomic parasites" (Yoder, Walsh, and Bestor 1997), suggesting to have no useful function for the host and their only purpose is to multiply themselves in the host genome. However, more recent studies have shown that they have a beneficial effect by creating genetic diversity in the genome (Nekrutenko and Li 2001), they influencing the host genomes and phenotypes (Beck et al. 2011; Hancks and Kazazian 2012).

DNA transposons and RNA transposons are the major types of mobile elements (Finnegan 1989). The present thesis focuses on RNA transposons, also called retrotransposons, which accumulate in the genome and influence the genomic composition.

SVA (SINE-R-VNTR-Alu, Shen et al. 1994) and LAVA (LINE1-Alu-VNTR-Alu, Carbone et al. 2012), together with PVA (PTGR2-VNTR-Alu, Hara et al. 2012) and FVA (FRAM-VNTR-Alu, Ianc et al. 2014), are part of the VNTR (or Variable Number of Tandem Repeats)-composite retrotransposons. They are specific to hominoid primates (Wang et al. 2005; Carbone et al. 2012; Hara et al. 2012; Ianc et al. 2014; Carbone et al. 2014). The mobilization of these elements involves proteins encoded by L1 elements (Ostertag et al. 2003; Hancks et al. 2011; Raiz et al. 2012; Carbone et al. 2014; Ianc et al. 2014). However, the details of their mobilization mechanism are not elucidated to date.

In this PhD thesis, the structural requirements in the mobilization of VNTR-composite elements were investigated. Special interest was paid in those structural / sequence features of human SVA and gibbon LAVA elements, that allow them to be efficiently mobilized.

1.2. Classification and general structure of mobile elements

Based on their mobilization process, mobile elements can be classified into: RNA transposons (class I) and DNA transposons (class II) (Finnegan 1989).

DNA transposons are mobilized via a DNA intermediate by "cut-and-paste" mechanism, also called transposition. The transposons are flanked by inverted repeats and encode a transposase which helps the element in self-excision (Smit and Riggs 1996) and after a target site cleavage, the insertion of the element from one genomic place to another.

Retrotransposons are mobilized by a "copy-and-paste" mechanism, via an RNA intermediate (Cordaux and Batzer 2009). Usually, their mobilization mechanism involves several

steps: transcription of the element (and translation in case of autonomous retrotransposons), reverse transcription into cDNA and next integration of cDNA copy in a new genomic location after the genomic target site cleavage. Thus, at the end of this process, an active RNA transposon is multiplied into the genome. As a mark of their mobilization, retrotransposons are flanked by TSDs and ended with a poly-A tail, also retrotransposons can be truncated or inverted and can transduce genomic sequences during retrotransposition process (Ostertag et al. 2003). They can be further classified based on the mobilization mechanism, as autonomous, which encode factors necessary for their own mobilization, and non-autonomous, which use the mobilization machinery of the autonomous retrotransposons (Ostertag and Kazazian Jr 2001).

1.2.1. Autonomous LINE1 element

Long INterspersed Element (LINEs) is the only active autonomous retrotransposon in the human genome. It comprises ~21% of the genome (Lander et al. 2001).

A full-length element (of about 6 kb in length) has a 5' UTR that contains an endogenous RNA polymerase II promoter (Swergold 1990), two open reading frames separated by a spacer and a 3’ UTR which contains a polyadenylation signal (Dombroski et al. 1991). The ORF1 encodes a nucleic acid-binding protein (Hohjoh and Singer 1996) with chaperone activity (Martin and Bushman 2001). The ORF2 encodes ORF2p with endonuclease and reverse transcriptase activity (Feng et al. 1996; Mathias et al. 1991).

1.2.2. The non-autonomous retrotransposons

The non-autonomous retrotransposons do not encode any proteins and their mobilization is possible only with proteins of autonomous elements. SINEs elements, VNTR-composite elements and processed pseudogenes fall into this category.

SINE (Short INterspersed Elements) are the most abundant class of retrotransposons in the human genome. Alu element, the main representant in this category, is a primates specific SINE element (Shen, Batzer, and Deininger 1991). Structurally, Alu elements (of ~300 bp in length) and are composed of two monomers, the left and the right monomer, originating from 7SL RNA (Ullu and Tschudi 1984) and contains an internal Pol III-promoter.

VNTR-composite elements are composed, as the name suggests, by different domains around of the central VNTR (variable-number-of-tandem-repeats) region. This category includes SVA (Shen et al. 1994), PVA (Hara et al. 2012), FVA (Ianc et al. 2014) and LAVA (Carbone et al. 2012) elements. They share a similar 5’ and central region and differs by their 3’ part. The 3’ end of the VNTR-composite elements was acquired by splicing (Ianc et al. 2014).

SVA element

The SVA element has ~2 kb and is composed at its 5' end by a CT-hexameric-repeat region (CCCTCT) and an Alu-like region which contains two antisense Alu fragments and a sequence with unknown origin; follows a variable number of 30-50 bp tandem repeats (VNTR) region (Zhu et al. 1992) and a SINE-R region (490 bp) at the 3' end. This SINE-R region has the origin in the 3' end of the *env* gene and the 3' part of LTR of the HERV-K10 (Ono, Kawakami and Takezawa 1987). The element ends with a canonical polyadenylation signal (AATAAA) and a poly-A tail (Wang et al. 2005).

SVA elements are transcribed by RNA polymerase II (Wang et al. 2005), like L1 elements (Swergold 1990), but the endogenous *pol II* promoter has not been identified to date.

SVAs are hominoid-specific elements, no SVA was found in the Old World monkeys (including the rhesus macaque). Tracking back the origin of SVAs, in the genome of the rhesus macaque, a precursor of the VNTR region present in SVA elements was identified. This precursor was also identified in the human genome (Han et al. 2007) and it was called SVA2 (it is formed by a VNTR region, followed by unique 3' region and a poly-A tail). So, from an evolutionary point of view, it can be said that SVA originates before the divergence of hominoid species, and after the divergence of Old World primates (Wang et al. 2005).

However, they were successfully amplified only in Great Apes (hominids), being very abundant in human (~2700) (Wang et al. 2005), chimp (~2600), gorilla (~2300) (Levy et al. 2017) and orangutan (~1800) (Locke et al. 2011) compared to gibbon which has only 29 copies of SVA (Ianc et al. 2014).

The hominoid SVA elements are divided into six major subfamilies based on diagnostic mutations compared to HERV-K10 sequence: named with A to F, from older to younger ones.

Being an active element, SVA cause inter-individual variation by the presence/absence polymorphism in humans (Wang et al. 2005; Bennett et al. 2004; Feusier et al. 2018). Also, it is known that SVA insertions influence the genomic landscape, by being a source of genetic diversity (Damert et al. 2009; Hancks et al. 2009; Quinn and Bubb 2014) but also can produce diseases in human (Kobayashi et al. 1998; Ostertag et al. 2003; Callinan and Batzer 2006).

LAVA element

LAVA element was discovered in the gibbon genome (Carbone et al. 2012). As all VNTR-composite elements, LAVAs contains a CT-hexameric region, followed by a sequence with homology to Alu elements, a VNTR central domain and ends by a region which originates from intron 2 of hydroxysteroid (17-beta) dehydrogenase 3 gene – this portion of intron 2 contains sequences from the AluSz and L1ME5 elements.

A total of 1797 LAVA copies were identified in the gibbon genome, classified into six larger families, named from LAVA_A to LAVA_F (Carbone et al. 2014). The older family considered a fossil element and possessing the longest Alu-like region, LAVA_A - presents high sequence similarity with Alu-like region of the old SVA_A family (Carbone et al. 2014; Ianc et al. 2014). The other LAVA families are more sequence-related to each other and contain specific Alu-like deletions (such as LAVA_F) or an insertion (such as LAVA_D) (Carbone et al. 2014; Ianc et al. 2014).

When analyzing the evolutionary history of LAVAs in primates, the LAVA prototype should be formed before the scission of gibbons and Great Apes. This assumption is based on the fact that 5' truncated LAVA elements – containing only the VNTR and LA part – were identified in the gorilla, chimpanzee and human genome (Damert 2018). No such intermediate was found the orangutan genome and nor in the Old World monkeys. Only the younger family, LAVA_F, seems to be amplified after the divergence of the four gibbon genera, and is *Nomascus leucogenys* specific (Carbone et al. 2014).

1.3. Mobilization of retrotransposons

1.3.1. Cis-mobilization of autonomous L1 element

The life cycle of L1 retrotransposition assumes that the genomic L1 element is transcribed by its internal promoter, followed by the transfer into the cytoplasm where the L1 mRNA is translated (McMillan and Singer 1993; Dmitriev et al. 2007). After translation, several ORF1p and some ORF2p together with a retrotransposition-competent L1 mRNA and other host factors, such as PABPC1/4, MOV10, UPF1 and ZCCHC3 (Taylor et al. 2018), are assembled into a ribonucleoprotein. In the next step, the L1-protein complex enters into the nucleus during mitosis, when the breaks down, and they remain trapped there when the new nuclear envelope forms (Mita et al. 2018). In the nucleus, the L1 RNA is reversely transcribed and integrated into a new genomic place by TPRT mechanism (Luan et al. 1993), mainly in the S phase of the cell cycle.

The TPRT begins once the antisense strand of the genomic target site (which is a T-rich sequence) was nicked by the endonuclease domain of the ORF2p, revealing a 3' hydroxyl group. Next, the poly-A tail of L1 RNA hybridizes to the TTTTs stretch of the nicked DNA, and the RT domain of ORF2p uses the exposed 3' OH end of the TTTTs to prime reverse transcription of the element into cDNA (Luan et al. 1993). Then, the second strand of the target site must be cleaved to create a primer for sense strand synthesis. The position of the second nick relative to the initial

one (downstream, upstream or in the same line with the first nick), will generate target-site-duplications (TSDs), which will flank the inserted element.

The cells have developed mechanisms to inhibit L1 mobilization, like DNA methylation, histone modification and piwi-interacting RNAs, mainly to maintain genomic integrity (Di'Giacomo et al. 2013; Yang and Wang 2016). In somatic cells, L1 mobilization is silenced by epigenetic mechanisms or post-transcriptional by siRNAs (Garcia-Perez et al. 2010; Yang and Kazazian 2006).

1.3.2. Trans-mobilization of non-autonomous retroelements

The non-autonomous retrotransposons, like Alu and VNTR-composite retroelements, do not encode proteins necessary for their mobilization and therefore they are mobilized in trans by L1 encoded machinery.

The success of Alu elements in the human genome is due to the hijacking of the L1 machinery - by miming the secondary structure of ribosome-associated RNAs. Thus, the Alu RNA might easily associate to ribosomes and reaches the proximity of the nascent L1 proteins (Dewannieux, Esnault and Heidmann 2003). Alu can in this mode to recruit the ORF2p as it is translated. This interaction seems to be determinant for Alu retrotransposition (Dewannieux, Esnault, and Heidmann 2003).

Until now, it was demonstrated the VNTR-composite elements can be mobilized in the human cell lines, in trans by L1 machinery (Ianc et al. 2014), and the dependency on L1 ORF1p (Hancks et al. 2011; Raiz et al. 2012; Ianc et al. 2014), but their whole mechanism of mobilization and host factors that are involved are waiting to be discovered. All studies regarding this subject were performed on SVA element. The critical role of 5’ end of the SVA in its efficient mobilization was evidenced through complete deletion of the 5’ end of the SVA_E element (the CT-hexameric repeats and its Alu-like region) which leads to a significant reduction of the retrotransposition rate (Raiz et al. 2012). An additional evidence was given by Hancks et al. who concluded that the 5’ end of SVA_D element represents “a minimal active SVA retrotransposon” (Hancks et al. 2012).

1.4. Aims of the thesis

VNTR-composite elements possess a similar structure but differ in their mobilization potential, as was evidenced by examining their copy number in the gibbon genome. Among VNTR-containing elements, SVA element had a great expansion in the human genome but not in the gibbon genome, where it exists in only 29 copies (Ianc et al. 2014). On the other hand, the gibbon genome is enriched in LAVA elements that are present in ~1700 copies (Ianc et al. 2014). In spite of the wide distribution of SVA and LAVA elements in the human and gibbon genomes, respectively, the molecular mechanisms of their propagation are poorly understood. In addition, the structural differences among VNTR-elements were not explored to date.

The main aim of this thesis was to explore the functional roles of SVA and LAVA domains in their mobilization process. Furthermore, the specific experimental objectives to aid in the achievement of the mentioned major aim were:

1. Investigation of structural domains and their structural features that mediate the efficient mobilization of SVA and LAVA. This objective is to be achieved by constructing SVA-LAVA chimeras through reciprocal domain exchange between the human active SVA_E element and two LAVA elements (*i.e.*, the active LAVA_F1 and inactive LAVA_E) and testing their retrotransposition potential.
2. Evaluation of the role of CT-hexameric region in the mobilization of VNTR-retrotransposons, particularly in SVA and LAVAs, by creating progressive 5' deletion mutants and testing their mobilization potential.
3. Elucidation of the role of the 3'end of the LAVA element in retrotransposition knowing that ~5 % of the LAVA elements of the *Nomascus leucogenys* genome are 3' truncated, downstream of the U2 region (Carbone et al. 2014; Ianc et al. 2014). This would be achieved also by progressive deletion of the 3' region and testing them in retrotransposition assay.
4. Assessment of the influence of CT repeats in nucleo-cytoplasmic traffic of SVA, by quantifying the SVA and SVAdeltaCT (SVA without the CT-hexameric region) RNA levels in nucleus and cytoplasm.

2. MATERIALS AND METHODS

2.1. Obtaining the SVA and LAVA's elements

The SVA_E element lacking their polyadenylation site was obtained from Annette Damert, cloned into pGEM Teasy and pCEPNeo.

LAVA elements from *Nomascus leucogenys* genome were chosen to belong to the younger LAVA subfamilies, LAVA_E and LAVA_F1 (Carbone at al. 2012; Carbone et al. 2014). LAVA elements were amplified by Annette Damert from gibbon genome (genomic DNA was provided by Christian Roos), from the chr3:108773434 - 108775518 (Nleu3.0) location in the case of LAVA_F and LAVA_E from chr2:155391066 - 155392835 (Nleu3.0). They were provided cloned into pJET1.2 and pCEPNeo, lacking their polyadenylation signals.

2.2. Testing the retrotransposition efficiency

2.2.1. Generation of SVA-LAVA domain-swaps

Chimeras were generated between SVA_E and LAVA_E/LAVA_F elements, by a reciprocal exchange between their 5'end (CT/Alu-like) and 3' end (VNTR/SINE-R; VNTR/LA).

The SL_E and L_ES chimeras contain domains combined from SVA_E and LAVA_E. First, the AlwNI/BstAPI sites of the SVA_E and LAVA_E were made compatible by amplification of the 5′ end (CT/Alu-like) using downstream mutagenetic primers. The obtained 5′ends of LAVA_E and SVA_E were next reciprocally combined with the SVA_E VNTR/SINE-R and LAVA_E VNTR/LA using BstAPI and AlwNI, into pCEPNeo.

SL_F and L_FS chimeras contain domains combined from SVA_E and LAVA_F. For generation of SL_F the exchange was made at the 3′ end. So, the SVA_E CT/Alu-like/VNTR was amplified using a mutagenetic downstream primer complementary to the SVA VNTR 3′-end which contains NcoI recognition situs (also present upstream of LAVA_F LA domain). The amplified SVA fragment was next combined with the LAVA_F LA-domain in pCEPNeo via KpnI/NcoI/NheI. The L_FS chimera was generated by amplification of the LAVA_F 5′ end using a mutagenetic primer to introduce a SmaI recognition site. The obtained fragment was next combined with the SVA_E VNTR/SINE-R into pCEPNeo using KpnI/(blunt)/NheI.

2.2.2. Generation of SVA and LAVA truncated elements

The CT-hexameric region of SVA and LAVA_F and also the 3'LA domain of LAVA_F were truncated. Progressive shortening of DNA fragments was performed with nuclease Bal-31 by progressive increasing the reaction incubation time. The resulting deletion mutants were

repaired at their ends with T4 DNA polymerase and then were ligated into the desired plasmid: pCEPNeo or pcDNA3deltaCMV.

2.2.3. Testing the mobilization potential of wild-type elements, chimeras and truncated elements

2.2.3.1. Transfections

4 x 10^5 Hela HA cells were seeded in T25 flasks 24 hours pre-transfection. The cells were co-transfected with 2 μg test plasmid and 2 μg pJM101L1RP_ΔNeo (L1-containing plasmid, provided by John Moran; the donor of L1 proteins required for retrotransposition), and using X-tremeGENE 9 DNA Transfection Reagent, according to the manufacturer's instructions.

Also, the pCEPNeo (empty vector) was tested for its ability to retrotranspose, as a control for the pseudogene formation rate. Wild-type elements were used for comparison.

2.2.3.2. Retrotransposition assay (trans-mobilization assay)

The mobilization potential of the SVA/LAVA chimeras, as well as that of the truncated SVA and LAVA elements, was tested using a cell-based retrotransposition assay. The trans-mobilization assay was performed as described by Moran et al. 1996 and Raiz et al. 2012, with some modifications presented in the thesis. The retrotransposition rates were expressed as the mean of G418-resistant colonies relative to that of SVA_E or to LAVA, which were considered 100%.

2.3. The influence of CT-repeats in nucleo-cytoplasmic traffic

2.3.1. Cloning the SVA and SVAdeltaCT into pcDNA3ΔCMV

To investigate whether the CT-hexameric repeat region of SVA has an influence on nucleo-cytoplasmic transport, SVA_E element containing or lacking the CT-domain (SVAdeltaCT) was cloned into pcDNA3ΔCMV plasmid (without its CMV promoter). In this mode, the SVA element will be expressed by its internal promoter. Elements were cloned via BglII(blunt)/NheI into pcDNA3ΔCMV (linearized with EcoRV/XbaI).

2.3.2. Transient transfections

Transfections were carried out on 1 x 10^6 HeLa HA cells, seeded in a 10 cm Petri dish, using 5 μg test plasmid and X-tremeGENE 9 DNA Transfection Reagent, according to the manufacturer's instructions.

2.3.3. RNA isolation and reverse transcription

24 hours post-transfection a confluent 10 cm Petri dish was split 1:2. After an additional 24 hours, the nuclear and cytoplasmic RNA was extracted.

First, the cell culture was washed with cold PBS three times and placed on ice. The cells were collected with a cell scraper and pelleted by centrifugation at 260xg for 5 minutes at 4˚C. Then the nuclear RNA was extracted according to Nevins protocol (Nevins 1987) and the cytoplasmic RNA was isolated based on solubilization of the plasma membrane with a hypotonic lysis buffer containing NP-40 (Farrell 2005 - protocol). The obtained RNA was stored at -80˚C.

Reverse transcription was carried out using the Verso cDNA Synthesis Kit, according to the manufacturer's instructions. Random Hexamers and Anchored Oligo-dT primers were used at a ratio of 3:1.

2.3.4. qPCR and data analysis

Specific primers and probe were designed to bind only SVAs expressed from the transfection plasmid, and to avoid amplification of the endogenous SVAs.

The expression of SVA was quantified using TaqMan probe detection, with the commercial kit Maxima Probe qPCR Master Mix. The commercial kit Maxima SYBR Green qPCR Master Mix was used to quantify the expression level of the endogenous PGK gene. All the primer sequences, reaction mixture and cycling conditions can be found in the thesis.

For data analysis the internal reference gene (PGK) was used, to determine fold-differences in the expression of the target gene (SVA and SVAdeltaCT).

The Livak method (Livak and Schmittgen 2001) was used for relative quantification of the SVA. The Ct was calculated automatically using Rotor-Gene 6000 Series Software 1.7. The Ct of the target gene (SVA, SVAdeltaCT) was normalized to that of the reference gene (PGK) in excel, obtaining the ΔC_t value. Next, the ΔC_t of the test sample was normalized to the ΔC_t of the calibrator, obtaining the $\Delta\Delta C$ value. And finally, was calculated the $2^{-\Delta\Delta Ct}$ ratio, expressed in fold increase/decrease of the target gene in the test sample relative to the calibrator sample. The Livak values were calculated independently for each transfection experiment, were averaged and plotted.

3. RESULTS

3.1. SVA and LAVA domains are incompatible

It was generated SL_E and L_ES chimeras, by reciprocal exchange of SVA and LAVA_E 5'ends. Based on the same idea, it was combined the domains of two active elements, like SVA_E and LAVA_F, to generate chimeras presumed to be active (SL_F and L_FS domain-swaps).

So, wild-type SVA_E, LAVA_E, LAVA_F elements, SVA-LAVA chimeras (SL_E, L_ES, SL_F, L_FS) and the pCEPNeo (empty vector) were all tested for their ability to retrotranspose in HeLa HA cells, in a trans-mobilization assay.

As it was expected, the wild-type LAVA_E had a low mobilization rate compared to SVA and to the LAVA_F –which showed again to be a very active element. In comparison to LAVA_E (which can be considered an inactive LAVA element), the SL_E chimera showed a minor increase in mobilization. This means that the 5' part of the SVA element could not enhance the LAVA_E mobilization. In the case of L_ES chimera, it was expected this result, it did not retrotranspose above pseudogene level (L_ES compared to pCEPNeo). The explication would be that L_ES has 5'CT-Alu-like domain from the inactive LAVA_E, which have a low mobilization potential.

Unexpectedly, the L_FS chimera, carrying the 5'CT-Alu-like from an active LAVA element, did not retrotranspose efficiently, nor the SL_F chimera; both being a combination of two active elements it would have been expected that the combination of domains from two active elements to give an active element. This result may indicate that SVA and LAVA mobilization pathways differ significantly.

Viewed as a whole, these results indicate that „SVA and LAVA domains are incompatible" and these two families of retrotransposons may have distinct mobilization pathways with different sequence/structural requirements. The data were published in Ianc et al. (2014). Further, the mobilization pathways remain to be elucidated.

3.2. CT-hexameric domain - essential for efficient mobilization

To test if the length of the CT-hexameric region can influence the mobilization efficiency, progressive 5' deletion mutants were generated for SVA_E and LAVA_F and they were next tested for their mobility, in a trans-mobilization assay. Five deletion mutants were obtained for SVA element and three for LAVA_F element, which covered the entire CT-hexameric region of these elements.

As I expected the retrotransposition results showed that all deletion mutants for SVA_E and LAVA_F1 are less efficiently mobilized than the full-length elements. In the case of SVA

mobilization, only a full-length CT-domain can support an efficient mobilization of the element. This result sustains again the existence of different mobilization requirements for SVA and LAVA elements.

3.3. LAVA 3'end has an inhibitory impact on mobilization

To evaluate the role of 3'region of LAVA element in retrotransposition, 3' truncations were achieved for LAVA_F element. Four representative 3' deletion mutants of LAVA_F were chosen. The selected clones are truncated in each region of the 3' LA part of the LAVA element.

The 3' truncated LAVA clones were tested in a retrotransposition assay for their mobilization potential. The retrotransposition results show that all of the 3'deletion mutants were more active than the full-length LAVA_F. This suggests that U1, AluSz and U2 regions are not crucial in the mobilization process. Surprisingly, the L1ME5 region, absent in all 3'deletion mutants and present only in the full-length LAVA element, showed to have an inhibitory impact on LAVA_F1 retrotransposition. The data were published in Ianc et al. (2014).

3.4. The influence of CT-hexameric repeats in the SVA nucleo - cytoplasmic transport

Little is known about SVA intracellular route. Only one study showed the subcellular SVA RNA distribution, being diffusely localized in the cytoplasm and in large aggregates in the nucleus (Goodier et al. 2010). But the mechanism of export into cytoplasm and re-import into the nucleus of SVA RNA remains unknown. Previous studies (Hancks et al. 2012; Raiz et al. 2012), but also the results of this thesis showed that the CT-hexameric repeats have an important role in the mobilization of SVA and LAVA. So, we next wanted to follow the CT-hexameric domain in nucleo-cytoplasmic transport. For this, a full-length SVA element and a 5'truncated SVA element, with the CT-hexameric region completly removed, were transiently transfected into HeLa HA cells and the SVA RNA (SVA cDNA) level from cytoplasm and nucleus was quantified, in order to see if the CT-hexameric region has a role in nuclear export and re-import.

The results obtained shows that in the nucleus both SVA and SVAdeltaCT are at higher levels than in the cytoplasm. The detected nuclear abundance of SVA and SVAdeltaCT confirms a previous study of Goodier et al. (2010). Also, the cellular distribution of transcribed SVA RNA seems to not be influenced by lack of the CT-repeats.

Further studies would be necessary to identify the exact cause of these obtained results. And also to identify the cellular factors that interact with the elements is mandatory in order to elucidate the SVA mobilization pathway.

4. DISCUSSIONS

Based on the results obtained in this thesis we could say that mobilization pathways of LAVA and SVA differ significantly. In this thesis, it was shown that the evolutionary youngest LAVA_F element is better mobilized than LAVA_E and than human SVA_E element. This difference in mobilization rate of LAVA_E and LAVA_F may be due to the different structure of the Alu-like region, the very active LAVA_F1 element having a shorter Alu-like region, most probably because of a splicing event, as a splice donor site could be detected at the truncation (Ianc et al. 2014). The differences among Alu-like sequences are also noticeable in the SVA subfamilies, the evolutionary younger SVAs also contains deletions in the Alu-like region. Furthermore, also the PVA, FVA are characterized by no deletions in the Alu-like domain and had lower retrotransposition rate (Ianc et al. 2014).

The modular structure of SVA and LAVA elements allowed testing of different domain combinations or different length elements in a trans mobilization assay to identify which domain makes them to be preferred by L1 machinery. We demonstrated that the 5' region is very important for efficient mobilization of SVA elements, but LAVA elements have different requirements for mobilization. We suggest the existence of two mobilization pathways for VNTR-composite elements. „SVA pathway" – in which the 5'end region is essential for an efficient mobilization, seems to be used also by the PVA and FVA elements (Ianc et al. 2014). However, for LAVAs the mobilization key seems to not being in the 5' end as in the case of SVA, PVA and FVA elements and nor in the 3' region. Lupan et al. (2015) showed that the VNTR region is a key determinant in LAVA mobilization, and CT-region seems to have a modulatory role in „LAVA pathway". The existence of LAVA pathway needs to be confirmed also for LAVA elements from others subfamilies and to be tested in a gibbon cell line in co-transfection with a gibbon L1 element.

The CT-hexameric region seems to resemble TOP motif - present in TOP mRNAs, a terminal oligopyrimidine sequence at their 5' end where the transcription initiation occurs by RNA pol II. It can be speculated that the observed role of CT-pyrimidine tract could be given by the proteins with which it could interact, like La or PTBP proteins. The La protein can bind the 5' terminal oligopyrimidine tract of TOP RNAs (Cardinali et al. 2003) and it can be presumed that La protein might be involved in retrotransposition. Also, PTBPs, which shuttle between nucleus and cytoplasm, are involved in splicing (Keppetipola et al. 2012) and interact with some RNAs that contain a sequence of 15-25 pyrimidines (Reid et al. 2009). This protein is involved in mRNA export and it was shown to be a co-factor in the nuclear export of hepatitis B virus

RNA (Zang et al. 2001). It would be interesting to investigate whether the La and PTBP proteins interact with VNTR composite RNAs.

Although the mobilization pathways of LAVA and SVA seems to differ significantly, these elements share some similarities in terms of genomic insertions. If explore the pre-existing LAVA and SVA insertions in the genome, it can be observed that about 63.4% of the SVA and 48% of the LAVA elements are full-length insertions (Wang et al. 2005; Carbone et al. 2014). Another similarity between SVA and LAVA insertions is given by the genomic insertion site, no preference for in or near genes insertion was observed (Hancks et al. 2011; Hancks et al. 2012; Raiz et al. 2012; Ianc et al. 2014). Regarding the intragenic insertion of LAVA elements, most of the pre-existent and *de* novo insertions occurred in introns (Ianc et al. 2014; Carbone et al. 2014). This low-frequency insertion in exons is common also for SVA copies in the human genome (Hancks et al. 2009).

5. CONCLUSIONS

This thesis is focused on the mobilization of two members of VNTR-composite retrotransposons - LAVA and SVA, and brings new information about the contribution of their domains in mobilization potential, offering a better understanding of the biology of VNTR-composite retrotransposons.

By using a combination of domains from SVA and LAVA elements were generated chimeras which were tested for their mobilization potential in order to have a first frame of what the structural requirements for retrotransposition are. We found out that chimeras derived from two very active elements (SVA and the younger LAVA_F element) were inactive. We concluded that SVA and LAVA domains are incompatible and most likely their mobilization pathways differ significantly, this being the first study that suggests the existence of two different mobilization pathways in VNTR-composite retrotransposons.

This is the first study showing the importance of CT-hexameric repeats in LAVA mobilization. We evidenced that all CT-deletion mutants for SVA_E and LAVA_F1 were less efficiently mobilized than the full-length elements, suggesting that CT-hexameric domain is crucial for the efficient mobilization of the VNTR-composite elements.

In addition, this thesis demonstrated for the first time the effect of the 3' end of LAVA elements on retrotransposition, all LAVAs 3' deletion mutants were better mobilized than the full-length LAVA_F1 element and we concluded that the L1ME5 region is responsible for the inhibitory effect on retrotransposition.

Finally, we attempted to further characterize the CT-hexameric region by investigating the influence of this region in nucleo-cytoplasmic traffic of SVAs. The results showed that the CT-hexameric domain might not have a significant influence on SVA nuclear transport. But, both full-length SVA and SVAdeltaCT, were found in a high quantity in the nucleus, confirming previous results that SVA forms aggregates into the nucleus (Goodier et al. 2010).

Identifying these structural domains that influence the mobilization of VNTR-composite elements is of a particular interest, because it could open up new research directions, like further investigation the host factors that interact with 5' and 3' domains and to elucidate the SVA and LAVA mobilization pathways. The results obtained in this thesis offered a start point for further studies performed by our lab colleagues, Lupan et al. (2015); they have investigated the role of the central VNTR-region in mobilization of LAVA elements.

REFERENCES

Beck CR., Garcia-Perez JL, Badge RM, and Moran JV. 2011. “LINE-1 Elements in Structural Variation and Disease.” *Annual Review of Genomics and Human Genetics* 12 (1): 187–215.

Bennett EA, Coleman LE, Tsui C, Pittard WS, and Devine SE. 2004. “Natural Genetic Variation Caused by Transposable Elements in Humans.” *Genetics* 168 (2): 933–51.

Callinan PA and Batzer MA. 2006. “Retrotransposable Elements and Human Disease.” *Genome Dynamics* 1: 104–15.

Carbone L, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, Meyer TJ, et al. 2014. “Gibbon Genome and the Fast Karyotype Evolution of Small Apes.” *Nature* 513 (7517): 195–201.

Carbone L, Harris RA, Mootnick AR, Milosavljevic A, Martin DIK, Rocchi M, Capozzi O, et al. 2012. “Centromere Remodeling in Hoolock Leuconedys (Hylobatidae) by a New Transposable Element Unique to the Gibbons.” *Genome Biology and Evolution* 4 (7): 648–58.

Cardinali B, Carissimi C, Gravina P, Pierandrei-Amaldi P. 2003. “La protein is associated with terminal oligopyrimidine mRNAs in actively translating polysomes.” *The Journal of biological chemistry* 278(37): 35145-51.

Cordaux R and Batzer MA. 2009. “The Impact of Retrotransposons on Human Genome Evolution.” *Nature Reviews Genetics* 10 (10): 691–703.

Damert A. 2018. “Lineage-Specific Evolution of SVA and LAVA Retrotransposons” *Molecular Life* 2 (1): 11–15.

Damert A, Raiz J, Horn AV, Löwer J, Wang H, Xing J, Batzer MA, Löwer R, Schumann GG. 2009. “5́-Transducing SVA Retrotransposon Groups Spread Efficiently throughout the Human Genome.” *Genome Research* 19 (11): 1992–2008.

Dawkins R. 1976. "The Selfish Gene." Oxford University Press.

Dewannieux M, Esnault C, Heidmann T. 2003. “LINE-Mediated Retrotransposition of Marked Alu Sequences.” *Nature Genetics* 35 (1): 41–48.

Di'Giacomo M, Comazzetto S, Saini H, De'Fazio S, Carrieri C, Morgan M, Vasiliauskaite L, Benes V, Enright A, O’Carroll D. 2013. “Multiple Epigenetic Mechanisms and the piRNA Pathway Enforce LINE1 Silencing during Adult Spermatogenesis.” *Molecular Cell* 50 (4): 601–8.

Dmitriev SE, Andreev DE, Terenin IM, Olovnikov IA, Prassolov VS, Merrick WC, Shatsky IN. 2007. “Efficient Translation Initiation Directed by the 900-Nucleotide-Long and GC-Rich 5’ Untranslated Region of the Human Retrotransposon LINE-1 mRNA Is Strictly Cap Dependent rather than Internal Ribosome Entry Site Mediated.” *Molecular and Cellular Biology* 27 (13): 4685–97.

Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH. 1991. “Isolation of an Active Human Transposable Element.” *Science* 254 (5039).

Farrell RE Jr. 2005. "RNA Methodologies: Laboratory Guide for Isolation and Characterization." *Elsevier/Academic Press*. 3th ed.

Feng Q, Moran JV, Kazazian HH, Boeke JD. 1996. “Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition.” *Cell* 87 (5): 905–16.

Feusier J, Watkins WS, Thomas J, Farrell A, Witherspoon DJ, Baird L, Ha H, Xing J, Jorde LB. 2018. “Pedigree-Based Estimation of Human Mobile Element Retrotransposition Rates.” *bioRxiv*, 506691.

Finnegan DJ. 1989. “Eukaryotic Transposable Elements and Genome Evolution.” *Trends in Genetics :* 5 (4): 103–7.

Garcia-Perez JL, Morell M, Scheys JO, Kulpa DA, Morell S, Carter CC, Hammer GD, et al. 2010. “Epigenetic Silencing of Engineered L1 Retrotransposition Events in Human Embryonic Carcinoma Cells.” *Nature* 466 (7307): 769–73.

Goodier JL, Mandal PK, Zhang L, Kazazian HH Jr. 2010. “Discrete Subcellular Partitioning of Human Retrotransposon RNAs despite a Common Mechanism of Genome Insertion.” *Human Molecular Genetics* 19 (9): 1712–25.

Han K, Konkel MK, Xing J, Wang H, Lee j, Meyer TJ, Huang CT, et al. 2007. “Mobile DNA in Old World Monkeys: A Glimpse Through the Rhesus Macaque Genome.” *Science* 316 (5822): 238–40.

Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH Jr. 2009. “Exon-Trapping Mediated by the Human Retrotransposon SVA.” *Genome Research* 19 (11): 1983–91.

Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH Jr. 2011. “Retrotransposition of Marked SVA Elements by Human L1s in Cultured Cells.” *Human Molecular Genetics* 20 (17): 3386–3400.

Hancks DC and Kazazian HH Jr. 2012. “Active Human Retrotransposons: Variation and Disease.” *Current Opinion in Genetics & Development* 22 (3): 191–203.

Hancks DC, Mandal PK, Cheung LE, Kazazian HH Jr. 2012. “The Minimal Active Human SVA Retrotransposon Requires Only the 5’-hexamer and Alu-like Domains.” *Molecular and Cellular Biology* 32 (22): 4718–26.

Hara T, Hirai Y, Baicharoen S, Hayakawa T, Hirai H, Koga A. 2012. “A Novel Composite Retrotransposon Derived from or Generated Independently of the SVA (SINE/VNTR/Alu) Transposon Has Undergone Proliferation in Gibbon Genomes.” *Genes & Genetic Systems* 87 (3): 181–90.

Hohjoh H and Singer MF. 1996. “Cytoplasmic Ribonucleoprotein Complexes Containing Human LINE-1 Protein and RNA.” *The EMBO Journal* 15 (3): 630–39.

Ianc B, Ochis C, Persch R, Popescu O, Damert A. 2014. “Hominoid Composite Non-LTR Retrotransposons-Variety, Assembly, Evolution, and Structural Determinants of Mobilization.” *Molecular Biology and Evolution* 31 (11): 2847–64.

Keppetipola N, Sharma S, Li Q, Black DL. 2012. “Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2.” *Critical Reviews in Biochemistry and Molecular Biology* 47(4): 360-378.

Kobayashi K, Nakahori Y, Miyake M, Matsumura K, Kondo-Iida E, Nomura Y, Segawa M, et al. 1998. “An Ancient Retrotransposal Insertion Causes Fukuyama-Type Congenital Muscular Dystrophy.” *Nature* 394 (6691): 388–92.

Kulpa DA and Moran JV. 2006. “Cis-Preferential LINE-1 Reverse Transcriptase Activity in Ribonucleoprotein Particles.” *Nature Structural & Molecular Biology* 13 (7): 655–60.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al. 2001. “Initial Sequencing and Analysis of the Human Genome.” *Nature* 409 (6822): 860–921.

Levy O, Knisbacher BA, Levanon EY, Havlin S. 2017. “Integrating Networks and Comparative Genomics Reveals Retroelement Proliferation Dynamics in Hominid Genomes.” *Science Advances* 3 (10): e1701256.

Livak KJ and Schmittgen TD. 2001. “Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method.” *Methods* 25 (4): 402–8.

Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, Yang SP, et al. 2011. “Comparative and Demographic Analysis of Orang-Utan Genomes.” *Nature* 469 (7331): 529–33.

Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 1993. “Reverse Transcription of R2Bm RNA Is Primed by a Nick at the Chromosomal Target Site: A Mechanism for Non-LTR Retrotransposition.” *Cell* 72 (4): 595–605.

Lupan I, Bulzu P, Popescu O, Damert A. 2015. “Lineage Specific Evolution of the VNTR Composite Retrotransposon Central Domain and Its Role in Retrotransposition of Gibbon LAVA Elements.” *BMC Genomics* 16 (1): 389.

Martin SL and Bushman FD. 2001. “Nucleic Acid Chaperone Activity of the ORF1 Protein from the Mouse LINE-1 Retrotransposon.” *Molecular and Cellular Biology* 21 (2): 467–75.

Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, Gabriel A. 1991. “Reverse Transcriptase Encoded by a Human Transposable Element.” *Science* 254 (5039): 1808–10.

McClintock B. 1950. “The Origin and Behavior of Mutable Loci in Maize.” *Proceedings of the National Academy of Sciences of the United States of America* 36 (6):344–55.

McMillan JP and Singer MF. 1993. “Translation of the Human LINE-1 Element, LlHs.” *Proceedings of the National Academy of Sciences of the United States of America* 90(24): 11533–11537.

Mita P, Wudzinska A, Sun X, Andrade J, Nayak S, Kahler DJ, Badri S, et al. 2018. “LINE-1 Protein Localization and Functional Dynamics during the Cell Cycle.” *eLife* 7: e30058.

Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr, Shibata H, et al. 1996. “High Frequency Retrotransposition in Cultured Mammalian Cells.” *Cell* 87 (5): 917–27.

Nekrutenko A and Li WH. 2001. “Transposable Elements Are Found in a Large Number of Human Protein-Coding Genes.” *Trends in Genetics* 17 (11): 619–21.

Nevins JR. 1987. “[22] Isolation and analysis of nuclear RNA.” *Methods in Enzymology* 152: 234–41.

Ohno S. 1972. “So Much "junk" DNA in Our Genome.” *Brookhaven Symposia in Biology* 23:366-70.

Ostertag EM and Kazazian HH Jr. 2001. “Biology of Mammalian L1 Retrotransposons.” *Annual Review of Genetics* 35 (1): 501–38.

Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr. 2003. “SVA Elements Are Nonautonomous Retrotransposons That Cause Disease in Humans.” *American Journal of Human Genetics* 73 (6): 1444–51.

Quinn JP and Bubb VJ. 2014. “SVA Retrotransposons as Modulators of Gene Expression.” *Mobile Genetic Elements* 4: e32102.

Reid DC, Chang BL, Gunderson SI, Alpert L, Thompson WA, Fairbrother WG. 2009. “Next-generation SELEX identifies sequence and structural determinants of splicing factor binding in human pre-mRNA sequence.” *RNA* 15(12): 2385-97.

Raiz J, Damert A, Chira S, Held U, Klawitter S, Hamdorf M, Löwer J, Strätling WH, Löwer R, Schumann GG. 2012. “The Non-Autonomous Retrotransposon SVA Is Trans-Mobilized by the Human LINE-1 Protein Machinery.” *Nucleic Acids Research* 40 (4): 1666–83.

Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel AW, Carroll MC, Zipf WB, Yu CY. 1994. “Structure and Genetics of the Partially Duplicated Gene RP Located Immediately Upstream of the Complement C4A and the C4B Genes in the HLA Class III Region. Molecular Cloning, Exon-Intron Structure, Composite Retroposon, and Breakpoint of Gene Duplication.” *The Journal of Biological Chemistry* 269 (11): 8466–76.

Shen MR, Batzer MA and Deininger PL. 1991. “Evolution of the Master Alu Gene(s).” *Journal of Molecular Evolution* 33 (4): 311–20.

Smit AF and Riggs AD. 1996. “Tiggers and DNA Transposon Fossils in the Human Genome.” *Proceedings of the National Academy of Sciences of the United States of America* 93 (4): 1443–48.

Swergold GD. 1990. “Identification, Characterization, and Cell Specificity of a Human LINE-1 Promoter.” *Molecular and Cellular Biology* 10 (12): 6718–29.

Taylor MS, Altukhov I, Molloy KR, Mita P, Jiang H, Adney EM, Wudzinska A, et al. 2018. “Dissection of Affinity Captured LINE-1 Macromolecular Complexes.” *eLife* 7: e30094.

Ullu E and Tschudi C. 1984. “Alu Sequences Are Processed 7SL RNA Genes.” *Nature* 312 (5990): 171–72.

Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. 2005. “SVA Elements: A Hominid-Specific Retroposon Family.” *Journal of Molecular Biology* 354 (4): 994–1007.

Yang F and Wang PJ. 2016. “Multiple LINEs of Retrotransposon Silencing Mechanisms in the Mammalian Germline.” *Seminars in Cell & Developmental Biology* 59: 118–25.

Yang N and Kazazian HH Jr. 2006. “L1 Retrotransposition Is Suppressed by Endogenously Encoded Small Interfering RNAs in Human Cultured Cells.” *Nature Structural & Molecular Biology* 13 (9): 763–71.

Yoder JA, Walsh CP and Bestor TH. 1997. “Cytosine Methylation and the Ecology of Intragenomic Parasites.” *Trends in Genetics* 13 (8): 335–40.

Zang W-Q, Li B, Huang P-Y, Lai MMC, Yen TSB. 2001. “Role of Polypyrimidine Tract Binding Protein in the Function of the Hepatitis B Virus Posttranscriptional Regulatory Element.” *Journal of Virology* 75(22): 10779-10786.

Zhu ZB, Hsieh SL, Bentley DR, Campbell RD, Volanakis DE. 1992. “A Variable Number of Tandem Repeats Locus within the Human Complement C2 Gene Is Associated with a Retroposon Derived from a Human Endogenous Retrovirus.” *The Journal of Experimental Medicine* 175 (6): 1783–87.