

UNIVERSITATEA BABEȘ-BOLYAI
FACULTATEA DE BIOLOGIE ȘI GEOLOGIE

TEZĂ DE DOCTORAT
(Rezumat)

**Transducția Alu și Poliadenilarea Prematură –
Aspecte de Biologie a Retrotranspozonilor SVA**

Conducător științific:

Prof. Dr. POPESCU Octavian

Doctorand:

CHIRA Sergiu

CLUJ-NAPOCA

2011

Cuprins

Listă abrevieri.....	3
Introducere.....	4
Obiective.....	8
Proiectarea a două seturi de vectori pentru investigarea rolului transducției Alu în mobilizarea SVA.....	9
1. Construcția setului de vectori H10_3.....	9
2. Construcția setului de vectori H10_1.....	11
Caracterizarea elementelor SVA 3' trunchiate.....	16
1. Secvența genomului uman de referință cuprinde 98 de elemente SVA 3' trunchiate	16
2. Regiunea SINE-R folosește o gamă largă de situsuri de poliadenilare alternativă.....	19
2.1 <i>Situsurile de poliadenilare alternativă ale elementelor SVA din secvența genomului uman de referință.....</i>	<i>19</i>
2.2 <i>Situsurile poli(A) folosite pentru poliadenilarea transcriptelor SVA 3' trunchiate exprimate în testiculul uman.....</i>	<i>23</i>
2.2.1 <i>Analiza 3' RACE.....</i>	<i>23</i>
2.2.2 <i>Originea și afilierea la subfamilie a transcriptelor SVA 3' intacte.....</i>	<i>25</i>
2.2.3 <i>Exprimarea, caracteristicile, elementele sursă și semnalele poli(A) utilizate de transcriptele SVA 3' trunchiate.....</i>	<i>26</i>
Concluzii.....	31
Bibliografie.....	32

Cuvinte cheie: retrotranspozoni SVA, transducție Alu, vectori raportori de retrotranspoziție, elemente SVA 3' trunchiate, poliadenilare prematură, transcripte SVA, testicul uman

Listă abrevieri

A	adenină
ADN	acid dezoxiribonucleic
ADNc	ADN complementar
ARN	acid ribonucleic
ARNm	ARN mesager
BAC	cromozom bacterian artificial
C	citozină
G	guanină
HERV	retrovirus endogen uman
kb	kilo baze
pb	perechi de baze
PCR	reacția în lanț a polimerazei
LTR	repetiție terminală lungă
SINE	elemente nucleare scurte intercalate
T	timină
TPRT	revers transcriere amorsată de matrită
TSD	duplicare a secvenței țintă
U	uracil
VNTR	repetiție numerică variabilă în tandem

Introducere

Elementele transpozabile, descoperite în 1940, sunt de asemenea cunoscute și sub numele de „gene săritoare”, datorită capacității lor de a se mobiliza de la o locație genomică la alta. Până la 45% din genomul uman este reprezentat de elemente transpozabile. Pe baza mecanismului de transpoziție, ele se pot împărți în transpozoni ADN și retrotranspozoni (Cordaux și Batzer, 2009). Retrotranspozonii reprezintă cea mai numeroasă clasă de elemente mobile, ocupând până la 42% din genomul uman. Ei se mobilizează printr-un mecanism de „copiere și lipire”, utilizând un intermediar ARN. Elementele care codifică factorii necesari mobilizării se numesc autonome și sunt reprezentate în genomul uman în special de retrotranspozonii non-LTR L1. Retrotranspozonii neautonomi, care includ pseudogenele procesate, Alu și SVA, nu codifică proteine, motiv pentru care depind de factorii codificați de retrotranspozonii autonomi (Cordaux și Batzer, 2009).

Unii dintre cei mai recentți din punct de vedere evolutiv retrotranspozoni neautonomi sunt elementele SVA, care s-au extins în genomul primatelor hominide în ultimii 18-25 milioane de ani. În genomul uman există aproximativ 2800 de copii, reprezentând 0,1% din genomul nostru (Wang *et al.*, 2005).

Acronimul SVA a fost introdus de Shen și colaboratorii (Shen *et al.*, 1994) și este în mod curent folosit (Hancks and Kazazian, 2010) pentru a defini o familie de retrotranspozoni cu structură compusă, SVA însemnând SINE-VNTR-Alu. Regiunea SINE-R este derivată din HERV-K10 și măsoară aproximativ 500 pb (Figura 1) (Ono *et al.*, 1987). Regiunea VNTR este foarte variabilă între elementele SVA și este compusă din copii ale unei secvențe de 30-50 pb (Wang *et al.*, 2005). În amonte de regiunea VNTR există o regiune de aproximativ 370 pb constituită din trei secvențe înrudite cu Alu, de 246, 54 și 25 pb în orientare antisens, și o secvență nedefinită (Shen *et al.*, 1994). Elementul întreg, care măsoară aproximativ 3 kb în lungime, este asociat la capătul 5' cu o repetiție simplă a hexamerului (CCCTCT). Coda poli(A) urmează după semnalul poli(A) AATAAA, iar întregul element este flancat la ambele capete de TSD-uri (Figura 1) (Ostertag *et al.*, 2003; Wang *et al.*, 2005). Mutațiile diagnostice față de secvența HERV-K10 au fost folosite pentru a stabili o ordine ierarhică a subfamiliilor

acestor elemente. Pe baza acestui criteriu au fost descrise șase subfamilii notate de la A la F, dintre care E și F sunt specifice omului (Wang *et al.*, 2005).

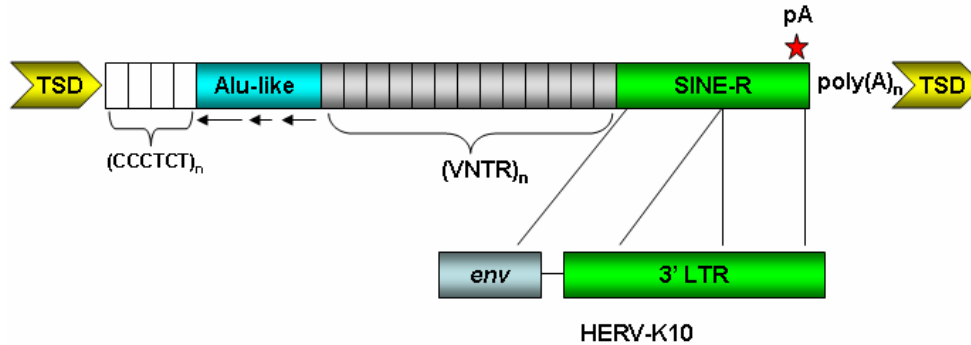


Figura 1. Structura compusă a retrotranspozonului neautonom SVA. Elementul cuprinde un număr variabil de repetiții al hexamerului CCCTCT, care este urmat de o regiune omoloagă cu secvențe Alu în orientare antisens, o regiune VNTR de lungime variabilă și o regiune derivată din HERV-K10 (SINE-R) care conține semnalul poli(A) (pA). Elementul se termină într-o coadă poli(A), iar întreaga structură este flancată de două TSD-uri (preluată din Ostertag *et al.*, 2003).

Cercetările menite pentru a identifica un promotor endogen care reglează transcrierea la SVA au condus la rezultate neconcludente (Damert *et al.*, nepublicat; Hancks *et al.*, 2009). În unele cazuri transcrierea este inițiată de promotori celulari din amonte, fapt care duce la transducția secvențelor de la 5' (Figura 2) (Damert *et al.*, 2009; Hancks *et al.*, 2009). De asemenea, folosirea unui semnal poli(A) din aval poate duce la transducția secvențelor genomice de la 3' (Figura 2) (Ostertag *et al.*, 2003; Xing *et al.*, 2006).

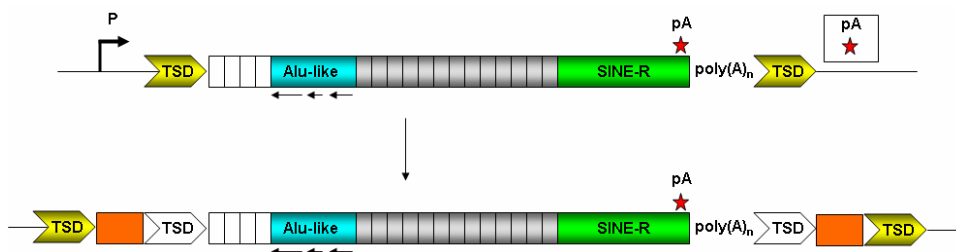


Figura 2. Comobilizarea de secvențe heteroloage prin fenomene de transducție la 5' și 3'. Transcrierea inițiată de la un promotor celular din amonte (P) duce la transducția secvenței 5' (boxa în portocaliu). Folosirea unui semnal poli(A) din regiunea din aval (boxa cu pA și stea roșie) duce la transducția regiunii 3' adiacente (boxa în portocaliu).

Printr-o analiză detaliată a cromozomului uman 19 a fost identificat un nou tip structural al elementelor SVA, caracterizat prin trunchierea capătului 3' a regiunii SINE-R. Aceste variante structurale se termină într-o coadă poli(A) și sunt flancate de două TSD-uri (Figura 3) (Damert *et al.*, 2009). Se presupune că acest tip de elemente au luat naștere prin poliadenilare prematură, dar până în prezent nu există date experimentale care să confirme această ipoteză.

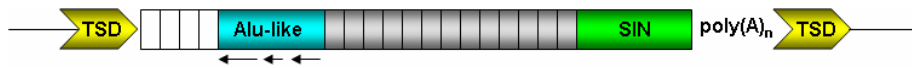


Figura 3. Structura unui element SVA trunchiat la capătul 3'. Elementul este caracterizat prin trunchierea la capătul 3' a regiunii SINE-R (SIN), terminându-se într-o coadă poli(A). Întreaga structură este flancată de două TSD-uri.

Ca și în cazul pseudogenelor procesate (Esanult *et al.*, 2000) și elementelor Alu (Dewannieux *et al.*, 2003), elementele SVA sunt de asemenea mobilizate *in trans* de către retrotranspozonii L1 (Hancks *et al.*, 2011; Raiz *et al.*, in press). Se presupune că regiunea Alu în orientare antisens a ARN SVA ar putea să hibridizeze cu un ARN Alu care este ancorat la ribozom (Figura 4). În felul acesta ARN SVA se poziționează în apropierea proteinelor L1, care facilitează mobilizarea (Mills *et al.*, 2007).

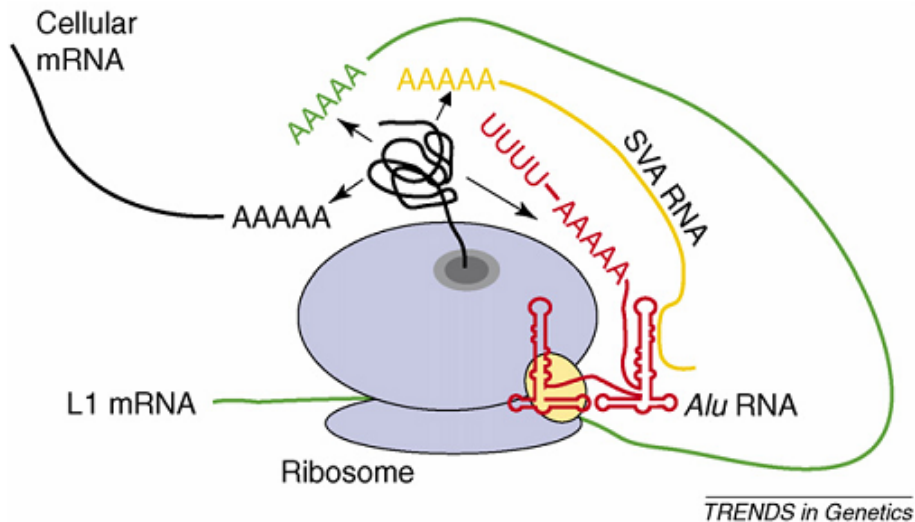


Figure 4. Modelul *trans*-mobilizării retrotranspozonilor neautonomi. Sunt ilustrate diferite scenarii pentru retrotranspoziția mediată de L1 *in cis* și *in trans*. Proteinele L1 traduse (linia neagră) interacționează cu ARN-ul care le-a codificat (linia verde), rezultând într-o preferință pentru mobilizarea *cis*. *Trans*-mobilizarea Alu implică ancorarea ARN Alu (linia roșie) la ribozom în apropierea proteinelor L1, „detunându-le”. Ancorarea la ribozom este mediată de structura secundară a ARN Alu, care facilitează legarea monomerului stâng de către particula de recunoaștere a semnalului, heterodimerul 9/14 (SRP9/14, ovalul galben). Se presupune că SVA hibridizează cu regiunea Alu-like la un RNA Alu ancorat la ribozom, poziționându-și ARN-ul în apropierea proteinelor L1. ARN-ul mesager celular (linia neagră) este ineficient mobilizat din cauză că nu este localizat în apropierea proteinelor L1. Săgețile negre indică interacțiunea dintre proteinele L1 și diferitele specii de ARN (modificat din Mills *et al.*, 2007).

Din motivul că unele inserții SVA au fost asociate în trecut cu diferite tipuri de afecțiuni umane (revizuite în Belancio et al., 2008), în ultimul deceniu a fost semnalat un interes crescând pentru aceste componente ai genomului nostru. Cu toate acestea, anumite aspecte legate de biologia lor au rămas neclarificate și trebuie adresate. Înțelegerea biologiei elementelor SVA poate contribui la o mai bună evaluare a impactului pe care acestea le-au avut și au în continuare asupra genomului nostru.

Scopul acestei teze este de a descrie contribuția autorului în domeniul cercetării retrotranspozonilor SVA.

Obiective

1. Elementele SVA sunt capabile să achiziționeze secvențe heteroloage prin fenomene de transducție la capetele 5' (Damert *et al.*, 2009; Hancks *et al.*, 2009) și 3' (Ostertag *et al.*, 2003; Xing *et al.*, 2006). Multe elementele SVA conțin elemente Alu integrale sau parțiale ca și secvențe transduse. Elementele Alu sunt eficient mobilizate prin interacțiunea cu heterodimerul SRP9/14 (Bennett *et al.*, 2008). În acest fel, elemente Alu integrale sau parțiale transduse ar putea crește eficiența mobilizării SVA, prin asigurarea structurii secundare necesară interacțiunii cu SRP9/14, sau, ca și în modelul lui Mills *et al.* (2007), prin hibridarea la un Alu ARN legat la SRP9/14. Pentru a dovedi o asemenea ipoteză sunt necesari vectori adecvați pentru testarea mobilizării SVA în culturi celulare.

Primul obiectiv al acestei teze este reprezentat de proiectarea a patru constructe SVA: un set de doi vectori care să conțină un element SVA cu sau fără un element Alu transdus în orientare sens, și un al doilea set de vectori care să conțină un element SVA cu sau fără un element Alu transdus în orientare antisens. Acești vectori pot fi utilizați în culturi celulare și rezultatele obținute din aceste experimente ar putea elucidă mecanismul *trans*-mobilizării elementelor SVA care conțin Alu.

2. Analiza elementelor SVA de pe cromozomul uman 19 a evidențiat un nou tip structural al acestor elemente, care este caracterizat prin trunchierea la capatul 3' a regiunii SINE-R (Damert *et al.*, 2009). Se presupune că aceste elemente ar fi luat naștere prin poliadenilarea prematură a ARN-ului derivat din elemente intacte la capătul 3'. O caracterizare detaliată a tuturor elementelor trunchiate la capatul 3' din genomul uman, complementată de identificarea de transcripțe SVA 3' trunchiate exprimate în testiculul uman ar putea evidenția faptul că poliadenilarea prematură este mecanismul responsabil pentru fenomenele de trunchiere la capătul 3' al elementelor SVA.

Proiectarea a două seturi de vectori pentru investigarea rolului transducției Alu în mobilizarea SVA

1. Construcția setului de vectori H10_3

Pentru a investiga o posibilă funcție a transducției Alu antisens în mobilizarea elementelor SVA, a fost selectat SVA H10_3 (Damert *et al.*, 2009). SVA H10_3 prezintă la capătul 5' un element AluSg pe lanțul minus și un element AluSx pe lanțul plus (Figure 5). AluSg în orientare antisens poate evidenția importanța elementelor Alu în mobilizarea SVA în concordanță cu modelul propus de Mills *et al.* (2007). În plus, datorită fenomenului de *splicing* observat între capatul 3' al AluSx și capătul 5' al AluSg în transducția 5' a H10_3 (Damert *et al.*, 2009), se poate obține o dovadă experimentală pentru ipoteza că regiunea Alu-like a elementelor SVA a luat naștere ca urmare a unui fenomen de *splicing* între elemente Alu (Hancks și Kazazian, 2010). Au fost construiți în acest scop un set de doi vectori, unul care conține SVA H10_3 cu secvențele AluSx/AluSg [Alu(+)] și unul care este lipsit de aceste elemente Alu [Alu(-)]. Pentru a asigura exprimarea SVA Alu(+/-) cu caseta raportoare *neo* (neomicină) a vectorului destinat pCEP-Neo (Raiz *et al.*, in press), semnalul poli(A) al elementului SVA H10_3 a fost înlăturat din vectorii de retrotranspoziție.

Locusul genomic conținând SVA H10_3 cu elementele AluSx/AluSg la 5' a fost amplificat prin PCR din clona BAC RPCIB753G12472Q (ImaGenes, Accession number AC073370), iar fragmentul de 4 kb obținut a fost clonat și secvențializat. Analiza datelor a evidențiat faptul că plasmidul S1 prezintă trei modificări în secvența de la capătul 5' (Figura 5).

Pentru a obține un vector de retrotranspoziție fără asemenea modificări, fragmentul 5' H10_3/Alu a fost reamplificat din ADN BAC, în timp ce fragmentul 3' H10_3 a fost reamplificat din plasmidul S1, apoi cele două fragmente au fost fuzionate, iar fragmentul întreg de 2.6 kb a fost clonat în pCEP-Neo între promotorul P CMV și caseta *neo*, obținându-se astfel vectorul de retrotranspoziție pSC1 (Figura 5).

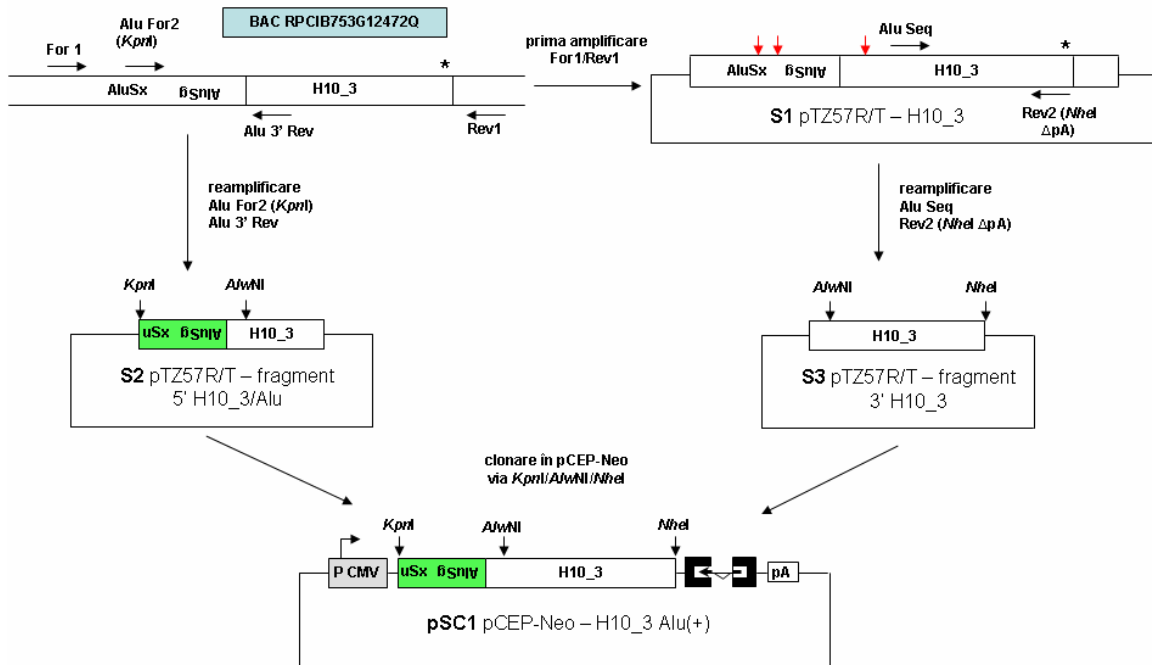


Figura 5. Construcția vectorului de retrotranspoziție SVA H10_3 Alu(+). Locusul genomic conținând SVA H10_3 AluSx/AluSg la capătul 5' a fost amplificat prin PCR din clona BAC RPCIB753G12472Q utilizând amorsele H10_3 For1 și Rev1, și clonat în pTZ57R/T, obținându-se plasmidul S1. Analiza prin secvențializare a evidențiat faptul că plasmidul S1 prezintă modificări în secvența clonată (săgeți roșii). Vectorul de retrotranspoziție H10_3 Alu(+) a fost construit prin reamplificarea regiunii 5' din BAC RPCIB753G12472Q cu amorsele H10_3 Alu For2 (*KpnI*) și Alu 3'REV. Regiunea 3' fără semnalul de poliadenilare (asterisce) a fost reamplificată din plasmidul S1 cu amorsele Alu Seq și H10_3 Rev2 (*NheI*, ΔpA). Ambele fragmente 5' și 3' au fost clonate via *KpnI/AlwNI/NheI* în pCEP-Neo, obținându-se vectorul de retrotranspoziție pSC1. P CMV – promotorul virusului citomegalic; boxa neagră cu săgeată inversată – caseta raportoare *neo*; ΔpA – deleția semnalului de poliadenilare.

Pentru a construi cel de-al doilea vector de retrotranspoziție, fragmentul H10_3 Alu(-) a fost reamplificat prin PCR din pSC1. Fragmentul de 1.8 kb a fost clonat în pCEP-Neo între promotorul P CMV și caseta *neo*, obținându-se astfel vectorul de retrotranspoziție pSC2 (Figura 6).

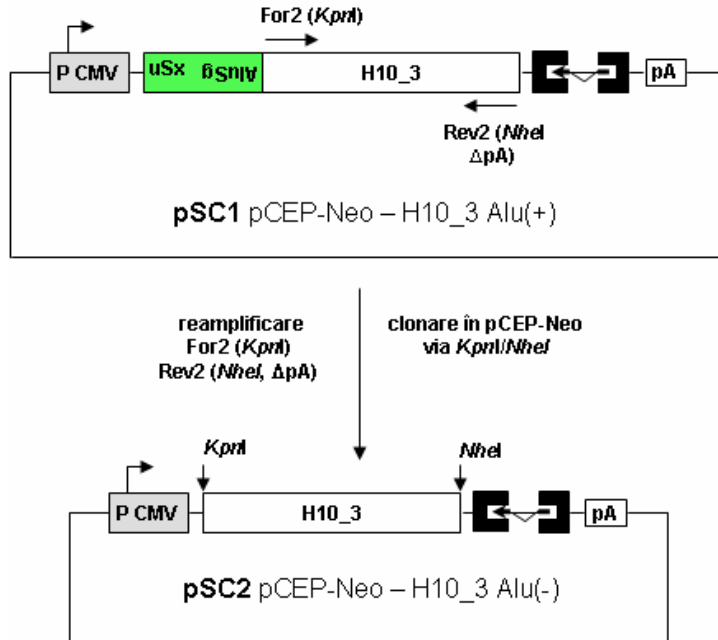


Figura 6. Construcția vectorului de retrotranspoziție SVA H10_3 Alu(-). Fragmentul SVA H10_3 Alu(-) a fost reamplificat din vectorul pSC1 cu amorsele H10_3 Alu For2 (*KpnI*) și H10_3 Rev2 (*NheI*, ΔpA) și clonat în pCEP-Neo, obținându-se vectorul de retrotranspoziție pSC2. P CMV – promotorul virusului citomegalic; boxa neagră cu săgeată inversată – caseta raportoare *neo*; ΔpA – deleția semnalului de poliadenilare.

2. Construcția setului de vectori H10_1

H10_1 este cel mai activ SVA identificat până în prezent, fiind elementul sursă pentru cel puțin 13 inserții din secvența genomului uman de referință (Damert *et al.*, 2009; Hancks *et al.*, 2009). SVA H10_1 prezintă un element AluSp la capătul 3' pe lanțul plus (Figura 7), care a fost transdus la toți cei 13 membrii ai grupului 4 din subfamilia F1 (Damert *et al.*, 2009). Acest fapt indică că elementele Alu în orientare sens transduse, ar putea mări eficiența mobilizării elementelor SVA care conțin elemente Alu, astfel că proiectarea unui set de vectori pentru culturi celulare care să conțină SVA H10_1 cu sau fără elementul AluSp de la capătul 3' [Alu(+)] ar putea furniza dovada pentru o asemenea ipoteza.

Pentru a asigura co-exprimarea elementului SVA H10_1 Alu(+/-) cu caseta raportoare *neo*, semnalele de poliadenilare de la capătul 3' al H10_1 și din secvența din aval de elementul AluSp (Figura 7) au fost înlăturate din vectorii de retrotranspoziție.

Locusul genomic care cuprinde SVA H10_1 cu elementul AluSp de la capătul 3' a fost amplificat prin PCR din clona BAC RPCIB753F0114Q (ImaGenes; Accesion number AL392107). Ampliconul de 4.2 kb a fost subclonat, obținându-se plasmidul S6. Analiza prin secvențializare a indicat că plasmidul S6 prezintă două deleții în coada poli(A) a elementului H10_1. Luând în considerare înlocuirea cozii poli(A) cu una sintetică (vezi mai jos), asemenea modificări nu se vor regăsi în vectorii de retrotranspoziție.

În plus față de această modificare irelevantă, un fragment de 835 pb din regiunea VNTR nu a putut fi secvențializată cu amorsele care flanchează regiunea VNTR. Pentru a secvențializa acest fragment, plasmidul S6 a fost digerat cu *AlwNI*, a cărui situs de recunoaștere este localizat în amonte de regiunea VNTR. Folosind exonucleaza Bal-31 și timpi de incubare crescători, au fost induse deleții progresive la capătul 5' a regiunii VNTR. Fragmentele rezultate au fost reparate la capete, clonate și secvențializate la capătul 5'. Prin această procedură întreaga regiune VNTR a putut fi secvențializată. Analiza datelor a evidențiat deleția unei G în regiunea VNTR. Cel mai probabil, această unică deleție nu afectează structura generală a elementului și în consecință plasmidul S6 a fost folosit pentru clonările ulterioare.

Elementul H10_1 fără AluSp a fost reamplificat din plasmidul S6, iar ampliconul de 3.5 kb a fost clonat, obținându-se plasmidul S7. Analizele prin secvențializare și cu enzime de restricție au indicat că fragmentul amplificat conține cinci modificări în secvența din amonte de situsul de recunoaștere a *BamHI* în plus față de o deleție de aproximativ 1 kb în regiunea VNTR (Figura 8).

Pentru a obține un fragment fără asemenea modificări semnificative, fragmentul cu mutații din plasmidul S7 a fost înlocuit cu fragmentul omolog din plasmidul S6, obținându-se în acest fel plasmidul S8. În final, elementul H10_1 Alu(-) a fost clonat în pCEP-Neo între promotorul P CMV și caseta *neo*, obținându-se astfel vectorul de retrotranspoziție pSC3 (Figura 8).

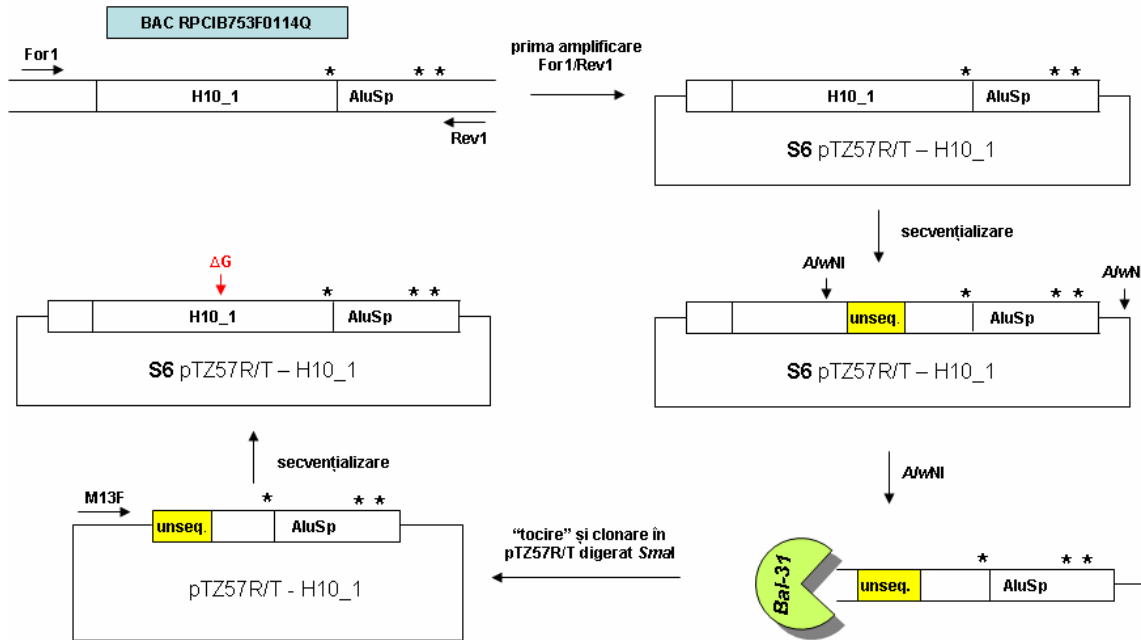


Figura 7. Amplificarea locusului genomic H10_1/AluSp și sevențializarea regiunii VNTR. Amorsele H10_1 For1 și Rev1 au fost folosite pentru a amplifica SVA H10_1 împreună cu elementul AluSp de la capătul 3' din clona BAC RPCIB753F0114Q și clonat în pTZ57R/T, obținându-se plasmidul S6. Analiza prin secvențializare a revelat că un fragment al regiunii VNTR a rămas nesevențializat (boxa galbenă). Pentru a secvențializa în întregime această regiune, plasmidul S6 a fost digerat cu *AwaNI*. Au fost create deleții progresive la capătul 5' al regiunii VNTR utilizând exonucleaza Bal-31. Fragmentele rezultate au fost „tocite” și clonate în pTZ57R/T digerat *SmaI*. Plasmidele rezultate, conținând regiunea VNTR de lungime variabilă [VNTR(-)_n], au fost secvențializate cu amorsa universală M13F. Analiza datelor de secvențializare a evidențiat faptul că plasmidul S6 prezintă o deleție a unui reziduu G în regiunea VNTR (ΔG cu săgeata roșie). Asterisc – semnal de poliadenilare.

Pentru a obține vectorul de retrotranspoziție H10_1 Alu(+), un fragment oligonucleotidic sintetic (Generi Biotech) a fost inserat în plasmidul S8, în aval de regiunea SINE-R a H10_1 pentru a substitui coada poli(A) a elementului H10_1 care conține semnalul poli(A), obținându-se plasmidul 10 (Figura 9). Separat, elementul AluSp a fost reamplificat din plasmidul S6, iar ampliconul de 389 pb a fost mai departe subclonat. Analiza prin secvențializare a indicat că plasmidul S11 are o deleție de două A în coada poli(A), în timp ce plasmidul S12 a avut o deleție G la capătul 5'. Asemenea modificări ar putea avea un impact asupra retrotranspoziției, deleția G putând afecta pliarea Alu ARN în conformația necesară interacțiunii cu SRP9/14 (Bennett *et al.*,

2008). De asemenea, pentru că lungimea cozii poli(A) joacă un rol esențial pentru mobilizarea elementelor Alu (Roy-Engel *et al.*, 2002; Dewannieux și Heidmann, 2005), asemenea modificări nu pot fi ignorate. Pentru a obține un element AluSp intact, fragmentul 5' al plasmidului S12 care conține deleția G a fost înlocuit cu fragmentul lipsit de mutații al plasmidului S11, obținându-se în acest fel plasmidul S13. Din acest plasmid, elementul AluSp a fost clonat în aval de coada poli(A) sintetică a H10_1 din plasmidul S10. Fragmentul H10_1 Alu(+) a fost inserat în pCEP-Neo între promotorul P CMV și caseta *neo*, obținându-se astfel vectorul de retrotranspoziție pSC4 (Figura 9).

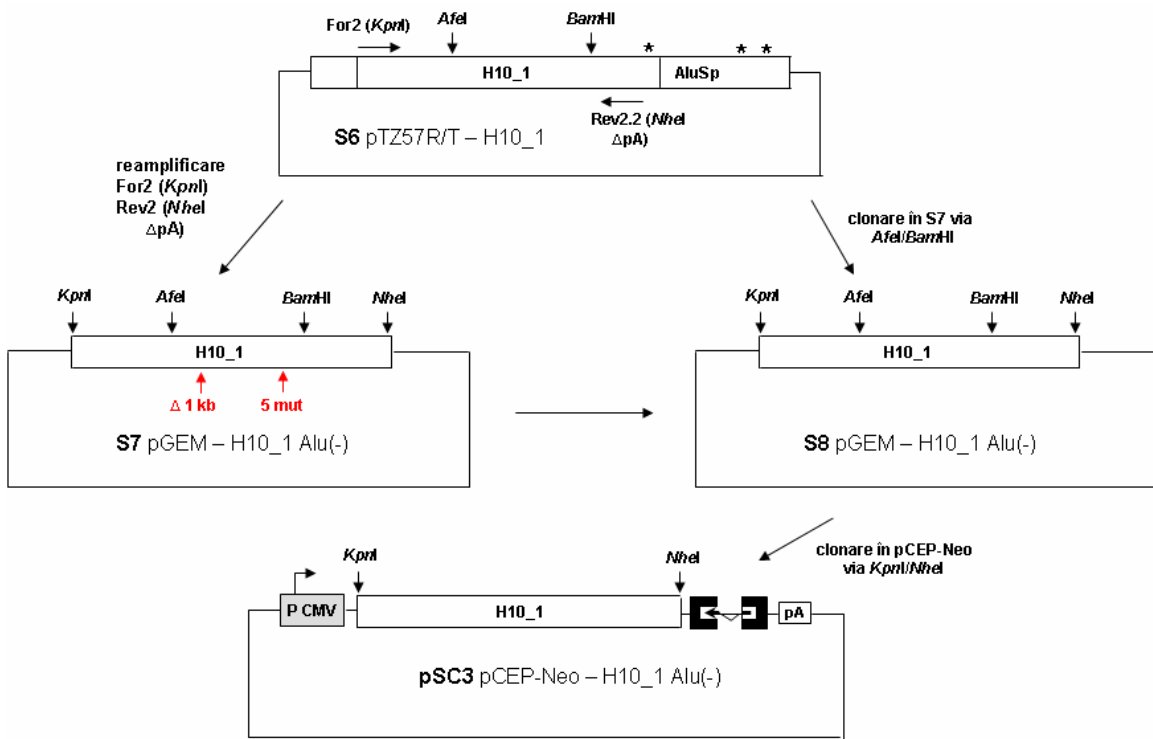


Figura 8. Construcția vectorului de retrotranspoziție SVA H10_1 Alu(-). Fragmentul H10_1 Alu(-) fără semnalul de poliadenilare (asterisc) a fost reamplificat din plasmidul S6 utilizând amorsele H10_1 For2 (*KpnI*) și H10_1 Rev2.2 (*NheI*, ΔpA) și clonat în pGEM, obținându-se plasmidul S7. Analiza prin secvențializare și restricție enzimatică a indicat că fragmentul H10_1 clonat prezintă cinci mutații la capatul 3' și o deleție de aproximativ 1 kb în regiunea VNTR. Pentru a obține un fragment lipsit de asemenea modificări, fragmentul *AfeI/BamHI* al plasmidului S7 a fost înlocuit cu fragmentul *AfeI/BamHI* al plasmidului S6, obținându-se în acest fel plasmidul S8. Vectorul de retrotranspoziție H10_1 Alu(-), pSC3 a fost generat prin clonarea fragmentului *KpnI/NheI* H10_1 Alu(-) în pCEP-Neo, între promotorul P CMV și caseta *neo* (boxa neagră cu săgeată inversată). ΔpA – deleția semnalului de poliadenilare.

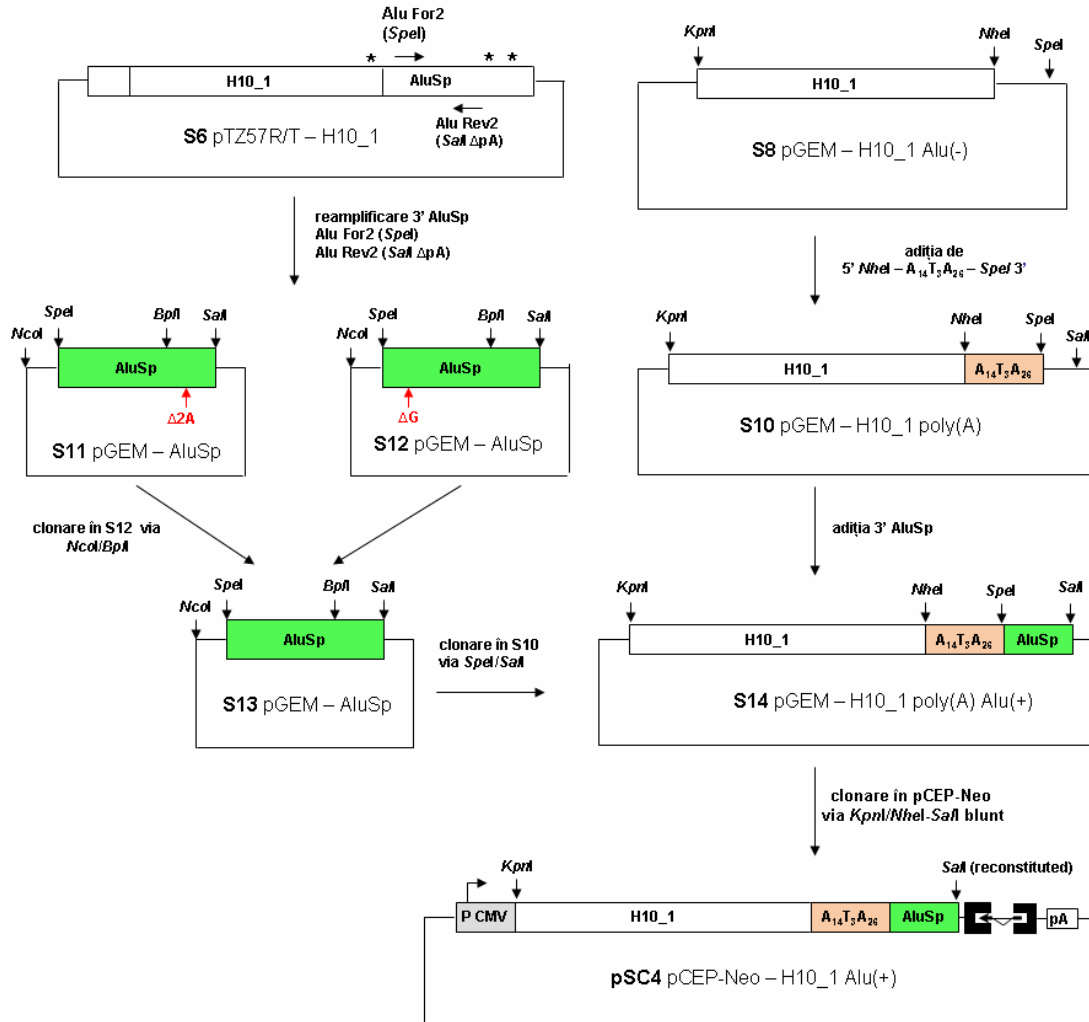


Figura 9. Construcția vectorului de retrotranspoziție SVA H10_1 Alu(+). Un fragment oligonucleotidic *NheI*-A₁₄T₃A₂₆-*SpeI* care substituie coada poli(A) ce conține semnalul de poliadenilare AATAAA, a fost inserat via *NheI/SpeI* la capătul 3' al H10_1 în plasmidul S8. Elementul AluSp a fost reamplificat fără semnalul de poliadenilare (asterisc) din plasmidul S6 utilizând amorsele H10_1 Alu For2 (*SpeI*) și H10_1 Alu Rev2 (*SalI*, ΔpA) și clonat în pGEM. Analiza prin secvențializare a două plasmide, a indicat că S11 prezenta două deleții de A în coada poli(A) a elementului AluSp (Δ2A cu săgeata roșie), în timp ce S12 conținea o deleție de G la capătul 5' al elementului AluSp (ΔG cu săgeata roșie). Un element AluSp intact a fost obținut prin înlocuirea fragmentului *NcoI/BpI* al plasmidului S12 cu fragmentul *NcoI/BpI* al plasmidului S11, obținându-se astfel plasmidul S13. Elementul AluSp reconstituit a fost clonat în plasmidul S10 via *SpeI/SalI*, în aval de fragmentul oligonucleotidic *NheI*-A₁₄T₃A₂₆-*SpeI*, obținându-se astfel plasmidul S14. Vectorul de retrotranspoziție H10_1 Alu(+), pSC4, a fost construit prin clonarea fragmentului H10_1/A₁₄T₃A₂₆/AluSp în pCEP-Neo via *KpnI/NheI-SalI* blunt, între promotorul P CMV și caseta *neo* (boxa neagră cu săgeată inversată). ΔpA – deleția semnalului de poliadenilare.

Caracterizarea elementelor SVA 3' trunchiate

1. Secvența genomului uman de referință cuprinde 98 de elemente SVA 3' trunchiate

Prin metoda de filtrare bazată pe adnotarea valorii de terminare a repetiției din Repeat Masker (<http://www.repeatmasker.org/cgi-bin/AnnotationRequest>) s-au identificat 104 elemente SVA în secvența genomului uman de referință HG17/2004. Analiza regiunii SINE-R a acestor SVA-uri a revelat că 98 elemente sunt trunchiate la capătul 3', trei elemente prezintă deleții interne în regiunea SINE-R, iar trei elemente au inserții de retrotranspozoni în regiunea SINE-R (Figura 10).

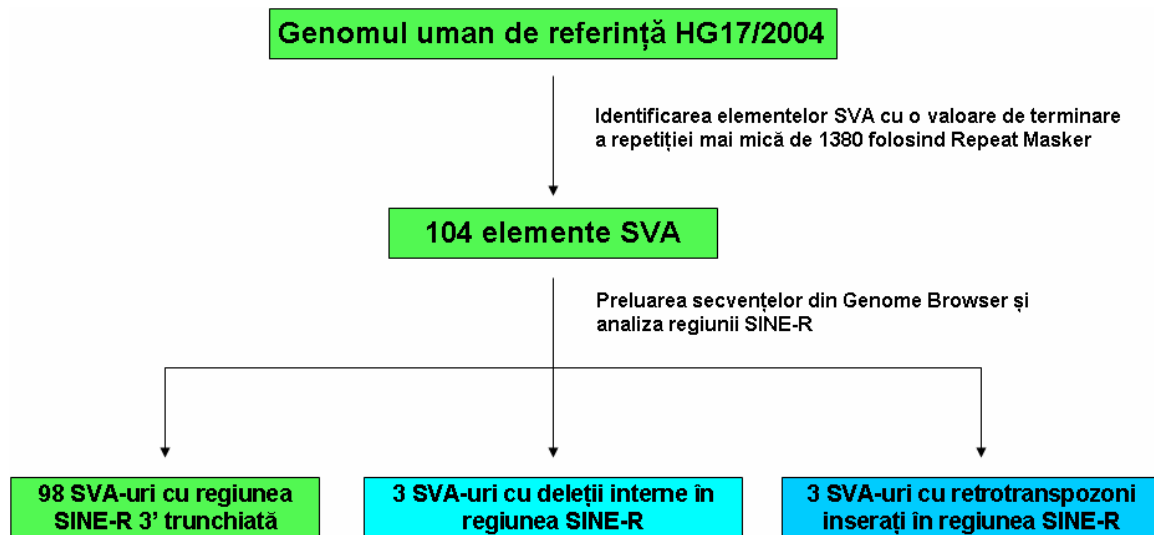


Figura 10. Schema identificării elementelor SVA 3' trunchiate. Într-o prima etapă secvența genomului uman de referință HG17/2004 a fost scanată pentru elemente SVA utilizând Repeat Masker. Cele aproximativ 2700 de elemente SVA identificate în acest fel (Damert *et al.*, 2009) au fost analizate pentru detectarea de elemente SVA cu o valoare de terminare a repetiției mai mică de 1380, care corespunde elementelor SVA cu regiunea SINE-R integrală. În a doua etapă, pe baza analizei regiunii SINE-R, cele 104 elemente identificate au fost împărțite în SVA-uri cu regiunea SINE-R 3' trunchiată (98 elemente), SVA-uri cu deleții interne în regiunea SINE-R (trei elemente) și SVA-uri cu retrotranspozoni inserați în regiunea SINE-R (trei elemente).

Cele 98 elemente cu regiunea SINE-R trunchiată la capătul 3' reprezintă 3.6% din numărul total de aproximativ 2700 elemente SVA din setul analizat. Opt dintre ele,

localizate pe cromozomul 19, au fost descrise anterior (Damert *et al.*, 2007). Lungimea regiunii SINE-R la cele 98 elemente variază de la 26 la 387 pb, având o lungime medie de 210 pb (Figura 11). Aceste rezultate indică faptul că elementele SVA 3' trunchiate sunt capabile de retrotranspoziție, regiunea SINE-R nefiind absolut necesară pentru mobilizarea SVA. Totuși, în elementele L1 poliadenilarea prematură care dă naștere la transcripte trunchiate la capătul 3', atenuază mobilizarea L1 (Belancio *et al.*, 2003).

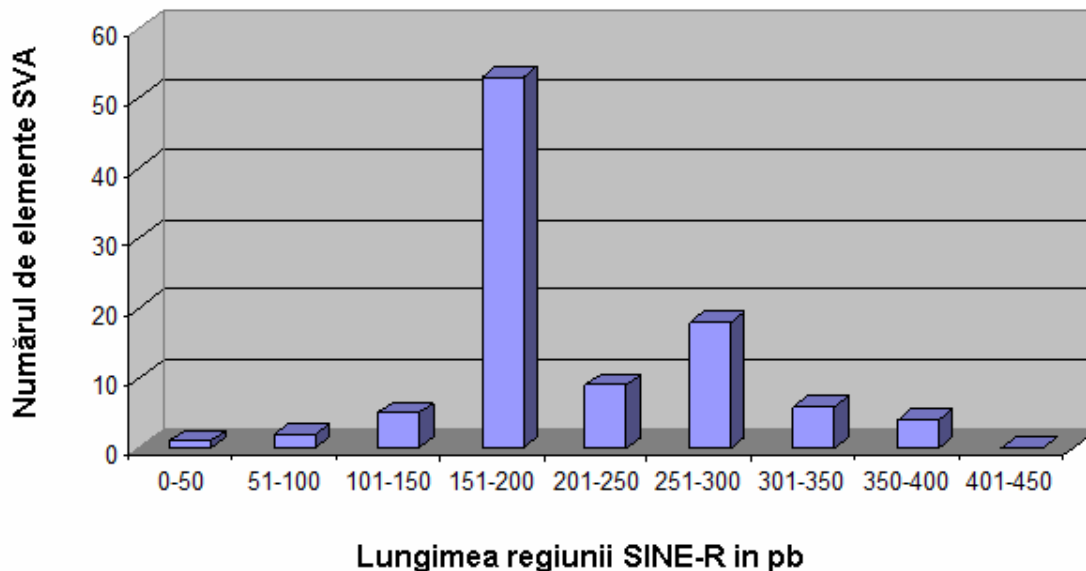


Figura 11. Distribuția lungimii regiunii SINE-R la elementele SVA 3' trunchiate. Abundența elementelor SVA 3' trunchiate în genomul uman (axa y) este indicată în corelație cu lungimea regiunii SINE-R, la intervale de 50 pb (axa x).

Lungimea cozilor poli(A) la elementele SVA 3' trunchiate variază de la 1 la 66 pb și sunt compuse fie numai din reziduuri de A, fie heterogene în compoziție. Această distribuție este similară cu distribuția elementelor SVA de pe cromozomul 19 (Damert *et al.*, 2009). În trei elemente, H1_69, H5_829 și H20_1877, potențiale cozi poli(A) coincid cu TSD-urile lor omogene în A sau bogate în A. Cozi poli(A) mai lungi ar fi putut exista în momentul inserției, dar s-au scurtat în timp (Chen *et al.*, 2005 și referințele cuprinse).

În două elemente, în care o coadă poli(A) lipsește la capătul 3' a regiunii SINE-R a fost observată o situație specială. În SVA H14_A378 și elementul lui descendent H2_413, regiunea SINE-R este fuzionată cu un fragment de 33 pb derivat din capătul 3'

al unui element L1PA. Un fenomen de splicing între ARN-ul SVA și L1PA este puțin probabil, pentru că regiunea SINE-R nu conține situsuri funcționale de splicing. O ipoteză ar fi schimbarea matriței în timpul TPRT (Gogvadze și Buzdin, 2005) între ARN L1PA și ARN SVA, ceea ce a rezultat în fuziunea capătului 3' al secvenței L1PA cu regiunea SINE-R a elementului SVA. În concluzie, aceste două elemente nu au luat naștere ca urmare a poliadenilării premature. O reprezentare de secvențe heteroloage fuzionate în SINE-R este ilustrată în figura 12.



Figura 12. Reprezentare schematică a regiunii SINE-R trunchiată la 3' (SIN în boxa verde) cu fuziuni de secvențe heteroloage (boxa de portocaliu închis). (A)_n – coadă poli(A).

În ceea ce privește afilierea la subfamilie a elementelor SVA 3' trunchiate, SVA_D este cea mai reprezentată, urmată de SVA_B și SVA_E. Această observație este în concordanță cu distribuția pe subfamilii a elementelor cu regiunea SINE-R integrală, dintr-un studiu amplu al elementelor SVA din genomul uman (Wang *et al.*, 2005). În schimb, subfamilia SVA_A este supra-reprezentată, ceea ce este în contrast cu distribuția elementelor SVA_A 3' intacte (Figura 13) (Wang *et al.*, 2005). O posibilă explicație pentru această diferență este aceea că elementele SVA_A ar putea furniza semnale poli(A) pentru poliadenilarea prematură (Wang *et al.*, 2005). Sau elementele afiliate acestei subfamilii au putut fi supuse unor fenomene de inserție/deleție/inversie mai intensiv comparativ cu elementele SVA afiliate la subfamiliile mai recente. Din acest motiv, elementele SVA_A ar trebui să fie supra-reprezentate printre cele 15% elemente SVA pentru care Wang și colaboratorii (Wang *et al.*, 2005) nu au stabilit o afiliere la subfamilie din cauza lipsei unei regiuni SINE-R intacte.

Aproximativ 55% din elementele SVA 3' trunchiate, afiliate la subfamiliile D, E și F, sunt inserții specifice omului, indicând faptul că mai mult de jumătate din elementele SVA 3' trunchiate s-au extins după divergența omului și cimpanzeului.

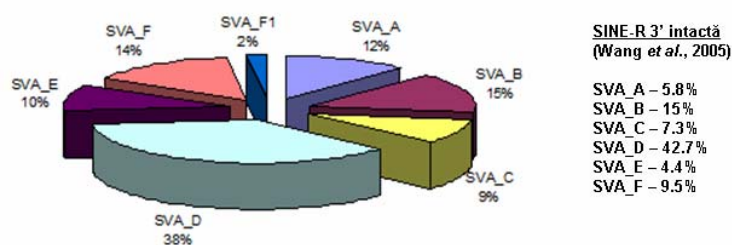


Figura 13. Distribuția pe subfamilii a elementelor SVA 3' trunchiate din genomul uman. Fiecare subfamilie este indicată ca și procent din numărul total de elemente SVA 3' trunchiate. Distribuția pe subfamilii a elementelor SVA cu regiunea 3' intactă este dată pentru comparație (Wang *et al.*, 2005).

2. Regiunea SINE-R folosește o gamă largă de situsuri de poliadenilare alternativă

2.1 Situsurile de poliadenilare alternativă ale elementelor SVA din secvența genomului uman de referință

Din această analiză au fost excluse cele două elemente care prezentau secvențe heteroloage fuzionate în regiunea SINE-R, din motivul că ele, cel mai probabil, nu au luat naștere prin poliadenilare prematură (vezi discuțiile de la pagina 17-18). În plus, un element SVA care conținea o secvență homopolimerică de C între regiunea SINE-R și coada poli(A) a fost exclus de asemenea din această analiză.

În funcție de poziția situsului de clivare și poliadenilare față de secvența consens SINE-R a subfamiliei, cele 95 de elemente rămase au fost împărțite în opt grupuri (Tabelul 1) și scanate pentru potențiale semnale poli(A) în amonte de situsul poli(A). În grupul 1 și 8, nu au putut fi identificate semnale de poliadenilare în secvența de 40 de baze din amonte de situsul de clivare și poliadenilare. Pentru restul grupurilor, potențialele semnale poli(A) identificate, reprezintă variante care diferă de semnalul AATAAA printr-o singură bază sau corespund secvenței consens NNUANA a semnalelor de poliadenilare umane (Beaudoing *et al.*, 2000). De asemenea, au fost identificate două semnale canonice, AATAAA și ATTAAA care au luat naștere prin substituția unei sigure baze în secvența de consens a subfamiliei (Tabelul 1).

În ceea ce privește distribuția semnalului poli(A) față de situsul de clivare și poliadenilare, acesta este localizat de la 4 la 37 perechi de baze în amonte față de situsul poli(A) (Figura 14). Într-un singur element, situsul de clivare și poliadenilare este

localizat imediat după semnalul poli(A), cum este de așteptat pentru elementele mobilizate de L1. Exceptând acest element, distribuția este similară cu poziția generală a semnalului poli(A), care se găsește la 10-35 baze în amonte de situsul de clivare și poliadenilare (revizuit în Millevoi and Vagner, 2010).

Tabel 1. Grupele de trunchiere cu potențialele semnale poli(A)

Grup	Poziția situsului de clivare și poliadenilare	Semnale poli(A) potențiale						
		SVA_A	SVA_B	SVA_C	SVA_D	SVA_E	SVA_F	SVA_F1
1	26						unknown (1)	
2	88-94		AAGAAA (1)	AATAGA GATAGA (1)				
3	132-135	TGTAGA AGTAGA (1)	TGTAGA AGTAGA (1)		TGTGGA GGTAGA (1)			
					TGTAGA AGTTGA (1)			
					TGTAGA AGTAGA (1)			
4	166-173	AAGAAA (2)	AAGAAA (8)	AAGAAA (5)	AAGAAA (17)	AAGAAA (7)	AAGAAA (6)	AAGAAA (1)
					AATAAA (1)			
5	194-205	TCTACA (1)	TCTATA (2)		TCTGTG (1)			
6	245-264	GTTAAA (1)	GTTAAA (1)		GTTAAA (5)	GTTAAA (1)	GTTAAA (2)	GTTAAA (1)
					GTTGAA (1)		ATTAAA (5)	
7	283-302	GTTAAA (5)	GTTAAA (3)	GTTAAA (2)	GTTAAA (3)	GTTAAA (2)		
8	362-385	unknown (1)		unknown (1)	unknown (2)			

Semnalele poli(A) potențiale sunt indicate în funcție de afilierea la subfamilie a fiecărui element

Numărul dintre paranteze indică numărul de elemente SVA care folosesc respectivul semnal(e) de poliadenilare

NNNNNN – semnale poli(A) potențiale care diferă printr-o singură bază față de secvența de consens a subfamiliei

NNNNNN – semnale poli(A) potențiale care diferă de NNUANA, dar corespund cu secvența de consens a subfamiliei

NNNNNN – semnale poli(A) potențiale care diferă printr-o singură bază față de NNUANA și secvența de consens a subfamiliei

Grup 1

A ACAGCTCCGAAAGAGACAGCGACCATCGAGAACGGCCATGATGACGATG
B ACAGCT-----CAATTGAGAACGGCCATGATGACGATG
C ACAGCT-----CAATTGAGAACGGCCATGATGACGATG
D ACAGCT-----CAATTGAGAACGGCCATGATGACAAATG
E GCGGCT-----CAATTGGGATGGCCATGATGACAAATG
F ACAGCT-----CAATTGAGAACGGCCAGGATGACAAATG
Consensus ACAGCT-----CAATTGAGAACGGCCATGATGACAAATG
1 33

Grup 2

A GTCGAAAAGAAAAGGGGAAATGTGGGAAAAGAAAGAGAGATCAGATTGTTACTGTG
B GTCGAATAGAAAAGGGGAAATGTGGGAAAAGAAAGAGAGATCAGATTGTTACTGTG
C GTCGAATAGAAAAGGGGAAATGTGGGAAAAGAAAGAGAAATCAGATTGTTACTGTG
D GTGGAATAGAAAAGGGGAAAGGTGGGAAAAGAAAGAGAAATCAGATTGTTACTGTG
E GTGGAATAGAAAAGGGGAAAGGTGGGAAAAGAAAGAGAAATCAGATTGTTACTGTG
F GTGGAATAGAAAAGGGGAAAGGTGGGAAAAGAAAGAGAAATCAGATTGTTACTGTG
Consensus GTCGAATAGAAAAGGGGAAAGGTGGGAAAAGAAAGAGAAATCAGATTGTTACTGTG
42 99

Grup 3

A TCTGTGTAGAAAAGTAGACATAGGAGAC---TC CATT TGT T
B TCTGTGTAGAAAAGTAGACATAGGAGAC---TC CATT TGT T
C TCTGTGTAGAAAAGTAGACATAGGAGAC---TC CATT TGT T
D TCTGTGTAGAAAAGTAGACATGGGAGACTTTT CATT TGT T
E TCTGTGTGATAGAAAGTAGACATGGGAGACTTTT CATT TGT T
F TCTGTGTAGAAAAGTAGACATGGGAGACTTTT CATT TGT T
Consensus TCTGTGTAGAAAAGTAGACATAGGAGACTTTT CATT TGT T
100 142

Grup 4

A TACTAAGAAAATTTCTTCTGCCTTGGGATCCTGT
B TACTAAGAAAATTTCTTCTGCCTTGGGATCCTGT
C TACTAAGAAAATTTCTTCTGCCTTGGGATCCTGT
D TACTAAGAAAATTTCTTCTGCCTTGGGATCCTGT
E TACTAAGAAAATTTCTTCTGCCTTGGGATCCTGT
F CACTAAGAAAATTTCTTCTGCCTTGGGATCCTGT
Consensus TACTAAGAAAATTTCTTCTGCCTTGGGATCCTGT
146 179

Grup 5

A TAAATCATAACCTTACCCCAACCCTGTCTCTCGA
B TAAATCATAACCTTACCCCAACCCTGTCTCTCGA
C TGATCTATGACCTTACCCCAACCCTGTCTCTCGA
D TGATCTGTGACCTTACCCCAACCCTGTCTCTCGA
E TGATCTGTGACCTTACCCCAACCCTGTCTCTCGA
F TGATCTGTGACCTTACCCCAACCCTGTCTCTCGA
Consensus TGATCTATGACCTTACCCCAACCCTGTCTCTCGA
180 216

Grup 6

A CAGGGTTAAATGGATTAAGGGCGGTGCAAGATG
B CAGGGTTAAATGGATTAAGGGCGGTGCAAGATG
C CAGGGTTAAATGGATTAAGGGCGGTGCAAGATG
D CAGGGTTAAATGGATTAAGGGCGGTGCAAGATG
E CAGGGTTAAATGGATTAAGGGCGGTGCAAGATG
F CAGGGTTAAATGGATTAAGGGCGGTGCAAGATG
Consensus CAGGGTTAAATGGATTAAGGGCGGTGCAAGATG
236 268

Grup 7

A CTTTGTAAACAGATGCTTGAAGGCAGCATGCTCGTT
B CTTTGTAAACAGATGCTTGAAGGCAGCATGCTCGTT
C CTTTGTAAACAGATGCTTGAAGGCAGCATGCTCGTT
D CTTTGTAAACAGATGCTTGAAGGCAGCATGCTCGTT
E CTTTGTAAACAGATGCTTGAAGGCAGCATGCTCGTT
F CTTTGTAAACAGATGCTTGAAGGCAGCATGCTCGTT
Consensus CTTTGTAAACAGATGCTTGAAGGCAGCATGCTCGTT
271 307

Grup 8

A GACACAAACACTGC GGAAGGCCGAGGGT CCTCT GCCTAGGAAAACAGAGACCTTTGTT
B GACACAAACACTGC GGAAGGCCGAGGGT CCTCT GCCTAGGAAAACAGAGACCTTTGTT
C GACACAAACACTGC GGAAGGCCGAGGGT CCTCT GCCTAGGAAAACAGAGACCTTTGTT
D GACACAAACACTGC GGAAGGCCGAGGGT CCTCT GCCTAGGAAAACAGAGACCTTTGTT
E GACACAAACACTGC GGAAGGCCGAGGGT CCTCT GCCTAGGAAAACAGAGACCTTTGTT
F GACACAAACACTGC GGAAGGCCGAGGGT CCTCT GCCTAGGAAAACAGAGACCTTTGTT
Consensus GACACAAACACTGC GGAAGGCCGAGGGT CCTCT GCCTAGGAAAACAGAGACCTTTGTT
345 404

(legenda pe pagina următoare)

Figura 14. Grupele de elemente SVA 3' trunchiate. Potențialele semnale poli(A) sunt subliniate în secvența de consens a subfamililor SVA_A-F. Pentru fiecare semnal, situsul de clivare și poliadenilare este indicat cu un capăt de săgeată, iar elementele SVA care utilizează acel situs de clivare și poliadenilare sunt indicate cu numere. Limitele secvenței, pentru fiecare grup, sunt indicate în secvența consens.

Aproximativ 50% dintre elementele SVA 3' trunchiate folosesc semnalul AAGAAA, în timp ce 25% folosesc semnalul de poliadenilare GTTAAA. 20% dintre elementele SVA 3' trunchiate folosesc semnale care corespund secvenței consens NNUANA sau care diferă printr-o bază de acest consens. Dar ce anume determină elementele SVA 3' trunchiate să folosească cu precădere semnalul AAGAAA? Secvențe potențitoare ar putea fi responsabile pentru această preferință.

Prin analiza secvenței de consens a grupului 4 au fost identificate două motive conservate TGTAG și TGTAC în amonte de semnalul AAGAAA (Figura 15). Acestea corespund secvenței consens de legare a factorului de clivare Im (UGUAN, CFIm) care este unul dintre factorii de poliadenilare. A fost demonstrat că CFIm poate direcționa poliadenilarea către semnalele poli(A) non-canonice prin intermediul factorului specific de clivare și poliadenilare (CPSF, Venkataraman *et al.*, 2005).

A doua secvență potențitoare a fost identificată ca fiind o secvență bogată în U localizată în amonte de semnalul AAGAAA (Figura 15). Secvențele bogate în U sunt capabile să direcționeze poliadenilarea către semnalele poli(A) non-canonice prin intermediul factorilor de *splicing* și a factorilor bazali ai poliadenilării (Danckwardt *et al.*, 2007). În acest fel, poliadenilarea la semnalul AAGAAA al grupului 4 ar putea fi influențată pozitiv de factori *splicing* și/sau de factori bazali de poliadenilare, prin interacțiunea cu secvențe potențitoare din amonte.



Figura 15. Elemente potențitoare din secvența de consens a grupului 4. Elementele *cis* sunt indicate în boxe colorate, iar factorii *trans* corespunzători sunt indicați în ovale colorate. CFIm – factorul de clivare Im. Literele subliniate – semnal poli(A).

În rezumat, aceste rezultate sugerează că poliadenilarea mediată de semnale poli(A) non-canonice este în mare măsură similară cu poliadenilarea transcriptelor genelor umane. Totuși, pentru un număr redus de elemente SVA, poliadenilarea pare a fi determinată de semnale poli(A) necunoscute. Există posibilitatea ca elementele SVA să folosească semnale poli(A) alternative total diferite. Aceasta poate indica că pozițiile conservate din secvența consens NNUANA (Beaudoing *et al.*, 2000) pot tolera mutații. Totuși, astfel de semnale reprezintă un substrat slab pentru interacțiunea cu factorii de poliadenilare, având în vedere numărul redus de elemente în care nu au putut fi detectate semnale poli(A).

2.2 *Situsurile poli(A) folosite pentru poliadenilarea transcriptelor SVA 3' trunchiate exprimate în testiculul uman*

2.2.1 *Analiza 3' RACE*

Metoda 3' RACE a fost folosită pentru a investiga dacă exprimarea transcriptelor SVA 3' trunchiate poate fi detectată în testiculul uman. Tehnica 3' RACE (Rapid Amplification of cDNA Ends – amplificarea rapidă a capetelor ADNc) cuprinde două etape și implică folosirea a trei amorse, una pentru sinteza ADNc și două pentru PCR (Figura 16). Amorsa pentru sinteza ADNc include o secvență ancoră la capătul 5' al unei secvențe oligo(dT). Această amorsă leagă ARNm poliadenilat care este revers-transcris în ADNc. În a doua etapă, capete 3' specifice sunt obținute din amestecul de ADNc heterogen prin PCR, utilizând o amorsă specifică și una care are doar secvența ancoră (Borson *et al.*, 1992).

Pentru a concepe o amorsă specifică pentru SINE-R, într-o primă etapă elementele SVA 3' trunchiate din genomul uman au fost sortate în 11 grupe pe baza valorii de terminare a repetiției din Repeat Masker (<http://www.repeatmasker.org/cgi-bin/AnnotationRequest>). Grupul 11 cuprinde numai un singur element. O secvență consens a fost construită pentru fiecare din primele 10 grupe, prin alinierea regiunii SINE-R a elementelor componente. Secvențele consens au fost aliniate cu secvența SINE-R a elementului grupului 11 pentru a obține o secvență consens finală în vederea proiectării amorsei (Figura 17). Amorsa astfel obținută, JS902+, are o substituție T/C în poziția 6 față de amorsa JS902 descrisă anterior (Kim *et al.*, 2000).

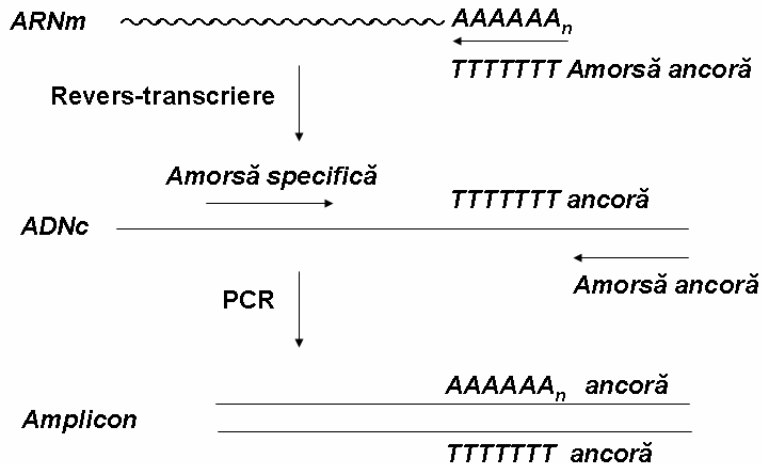


Figura 16. Principiul procedurii 3' RACE (**r**apid **a**mplification of **c**DNA **e**nds – **a**mplificarea rapidă a capetelor a ADNc). În prima etapă o amorsă oligo(dT)-ancoră este folosită pentru revers-transcrierea ARNm poliadenilat (ARNm) în ADNc. Acest ADNc este folosit mai departe ca și matriță pentru amplificarea capetelor 3', care includ și coada poli(A) (AAAAAA_n), utilizând o amorsă specifică și o amorsă ancoră.

```

gr1      --ACAGC--T CATTGAGAAC GGGCCATGAT GACAATGGCGGTTTT GTGGAATAGAAAAGG
gr2      --ACAGCGAT CATTGAGAAC GGGCCATGAT GACGATGGCGGTTTT GTCGANTAGAAAAGG
gr3      --ACAGC--T CATTGAGAA -GGGCCAGGAT GACAATGGCG -TTTT GTGGAATAGAAAAGG
gr4      --ACAGC--T CATTGAGAAC GGGCCAGGAT GACAATGGCGGCTTT GTGGAATAGAAAAGG
gr5      --ACAGC--T CATTGAGAAC GGGCCATGAT GACAATGGCGGTTTT GTGGAATAGAAAAGG
gr6      --ACAGC--T CATTGAGAAC GGGCCATGAT GACGATGGCGGTTTT GTCAAATAGAAAAGG
gr7      ACACAGC--T CATTGAGAAC GGGCCATGAT GACAATGGCGGTTTT GTGGAATAGAAAAGG
gr8      --ACAGC--T CATTGAGAGC GGGCCATGAT GACAGTGGCGGCTTT GTGGAAGTGGAGGGGG
gr9      --ACAGC--T CATTGAGAAC GGGCCATGAT GACAATGGCGGTTTT GTGGAATAGAAAAGG
gr10     --ACAGC--T C -TTGAGAAC GGGCCATGAT GACGATGGCG --TTTTGTAATAGAAAAGG
gr11     --ACAGC--T CATTGAGAAC GGGCCGGGATAAAAA -----AAAAA
Consensus --ACAGC--T CATTGAGAAC GGGCCATGAT GACAATGGCGGTTTT GTGGAATAGAAAAGG

```

JS902+ GAGAACGGGCCATGATGAC

Figura 16. Reprezentarea schematică a amorsei specifice SINE-R utilizată în 3' RACE. Secvențele grupurilor 1-10 reprezintă secvențele de consens care au fost construite prin alinierea regiunii SINE-R a elementelor din fiecare grup. Secvența grupului 11 reprezintă regiunea SINE-R a singurului element component. În verde este indicată secvența consens finală construită prin alinierea secvențelor consens a grupurilor 1-11. Amorsa JS902+ a fost concepută pe baza secvenței marcată în verde.

Amorsele JS902+ și Ancoră au fost utilizate pentru amplificarea SVA din ADNc din testicul uman. Rezultatele sunt ilustrate în figura 18. Ampliconul de aproximativ 520 pb indică faptul că secvențe SVA SINE-R integrale au fost amplificate din ADNc.

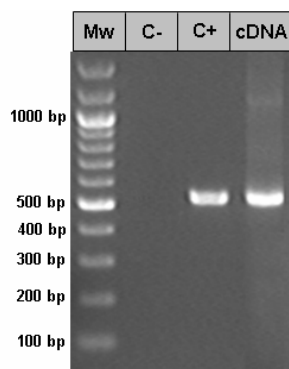


Figure 18. Produși de amplificare obținuți prin 3' RACE din ADNc din testicul uman. Controlul pozitiv (C+) reprezintă regiunea integrală SINE-R a elementului H10_1 (Damert *et al.*, 2009, Hancks *et al.*, 2009). C(-) – controlul negativ (fără ADNc); Mw – GeneRuler 100 bp.

2.2.2 Originea și afilierea la subfamilie a transcriptelor SVA 3' intacte

Prin clonarea și secvențializarea ampliconului de aproximativ 520 pb au fost obținute 14 secvențe SVA SINE-R care provin din zece elemente SVA sursă din genom (Tabelul 2). Șapte din aceste elemente SVA sunt membrii ai subfamililor SVA_A, B, C, iar restul de trei sunt afiliate la subfamilia SVA_D. Nici o secvență SVA_E sau F nu a fost identificată. Acest fapt nu exclude posibilitatea ca asemenea elemente sunt exprimate în testiculul uman. Elementele SVA_E și SVA_F reprezintă numai 4,4%, respectiv 9,5% din numărul total de elemente SVA din genomul uman (Wang *et al.*, 2005). Această abundență relativ mică comparativ cu celelalte subfamilii ar putea fi insuficientă pentru a putea fi detectate în numărul relativ mic de secvențe analizate.

Table 2. Transcripte SINE_R SVA 3' intacte exprimate în testiculul uman

Numele secvenței	Elementul sursă	Aflierea la subfamilie
IB1L1	H2_361	SVA_A
HB1L3	H5_858	SVA_A
IB1L3		
HB1L1	H13_A287	SVA_A
S11	H21_1992	SVA_A
HB1L4		
S3	H6_978	SVA_B
S2	H7_1225	SVA_B
S5	H7_1211	SVA_C
S16		
IB1L5		
S7	H1_179	SVA_D
S14	H5_873	SVA_D
S15	H5_859	SVA_D

2.2.3 *Exprimarea, caracteristicile, elementele sursă și semnalele poli(A) utilizate de transcriptele SVA 3' trunchiate*

O procedură „shotgun” a fost folosită din motivul că nu au putut fi identificați pe gel ampliconi care să corespundă elementelor SVA 3' trunchiate (Figura 18). Benzile de gel corespunzătoare unei lungimi de 400 pb, 300 pb și 200 pb pentru marker-ul ADN (Mw) au fost tăiate, purificate, iar produșii de purificare au fost clonați și secvențializați. Din numărul total de 21 de secvențe SVA obținute (Tabelul 3), o singură secvență SINE-R cu o lungime de 415 pb conținea o deleție internă de 84 pb.

Table 3. Transcripte SVA trunchiate în regiunea SINE_R, exprimate în testiculul uman

Numele secvenței	Lungime a SINE-R (pb)	Caracteristici	Element sursă 3' trunchiat/intact		Afilierea la subfamilie	Semnale poli(A) potențiale
IB3L4	133	3' trunchiat	H2_434	3' intact	SVA_B	TGTAGA/AGTAGA
IB3L10	83		H2_418		SVA_C	AATACA/GATAGA
IB4L1	171		H19_105		SVA_C	AAGAAA
IB3L6	171		H2_456		SVA_D	AAGAAA
IB2L10	171		H19_17		SVA_D	AAGAAA
IB2L1	73		H19_70		SVA_D	AATACA/GATAGA
IB4L10	261				GTTAAA	
IB3L3	220	3' trunchiat	H7_1026	3' trunchiat	SVA_A	TCTACA
IB3L5	149		H12_A190		SVA_A	TGTAGA/AGTAGA
IB3L7	149					
IB3L9	149					
IB2L9	253		H11_A52		SVA_D	GTTGAA
IB3L1	253					
IB4L5	171		H14_A405		SVA_D	AAGAAA
IB4L7	171		H16_A573		SVA_D	AAGAAA
IB3L2	170	3' trunchiat	nedeterminat	nedeterminat	nedeterminat	AAGAAA
IB3L8	173					
IB2L6	161	3' trunchiat; transducție la capătul 3'	H14_A413	3' trunchiat	SVA_C	AAGAAA derivat din transducția de la capătul 3'
IB2L2	161	3' trunchiat; fuziune la capătul 3' cu L1	H14_A378	3' trunchiat; fuziune la capătul 3' cu L1	SVA_D	AATAAA derivat din L1
IB2L4	161					
IB4L3	415	deleție internă în SINE-R	H20_1873	deleție internă în SINE-R; transducție la capătul 3'	SVA_D	AATAAA

Restul de 20 de secvențe SVA sunt trunchiate la capătul 3', cu lungimea SINE-R variind între 73 și 261 pb, având o lungime medie de 170 pb. Unul dintre ele prezintă

o transducție *de novo* la capătul 3', în timp ce două transcripte prezentau fuziuni cu secvențe L1PA (vezi discuțiile de la paginile 17-18).

18 din 20 de transcripte SVA 3' trunchiate provin din 13 elemente SVA sursă din genom, în timp ce la două transcripte nu s-a putut identifica elementul sursă (Tabelul 3). Analiza locilor de origine a evidențiat că șase din cele 13 elemente aveau regiunea SINE-R integrală. Restul de șapte elemente SVA sursă aveau regiunea SINE-R trunchiată la capătul 3'. În ceea ce privește afilierea la subfamilie, șapte sunt SVA_D, trei sunt SVA_C, două sunt SVA_A și un element este SVA_B. La fel ca și în cazul transcriptelor 3' intacte, nu au putut fi identificate elemente SVA afiliate la subfamiliile SVA_E și SVA_F.

Pentru a determina semnalele poli(A) care dirijează poliadenilarea la situsuri poli(A) alternative din regiunea SINE-R în transcriptele SVA 3' trunchiate exprimate în testiculul uman, cele trei transcripte care prezentau transducții și fuziuni la capătul 3' (Tabelul 3) au fost excluse din această analiză. Cel mai probabil, semnalul poli(A) al transducției/fuziunii a fost folosit pentru poliadenilarea acestor transcripte.

Pentru restul de 17 transcripte SVA clonate, potențialele semnale poli(A) identificate au fost categorizate în același mod ca și pentru grupele de trunchiere a elementelor din genomul uman. Ele sunt în principal reprezentate de semnale hexamerice care sunt potențial folosite și de către elementele SVA 3' trunchiate din secvența genomului uman de referință (Tabelul 4). Totuși, au fost identificate de asemenea noi potențiale semnale poli(A). Un asemenea semnal este AATACA care este folosit de un element SVA_C din grupul 5. Acest semnal a luat naștere ca urmare a substituției bazei G a semnalului AATAGA din secvența de consens a subfamiliei SVA_C. Al doilea semnal poli(A), GGTGAA, are o substituție G în poziția patru din secvența GGTA AAA.

Poliadenilarea dirijată de aceste potențiale semnale poli(A) (Tabelul 4) este în mare măsură similară cu cea a elementelor SVA 3' trunchiate din secvența genomului uman de referință. Poziția situsului de clivare și poliadenilare este aceeași pentru un semnal poli(A) specific. Numai în trei cazuri situsul de clivare și poliadenilare este poziționat diferit (Figura 18). În ambele cazuri, cel al elementelor SVA 3' trunchiate din secvența genomului uman de referință și al celor exprimate în testiculul uman,

situsul de clivare și poliadenilare și coada poli(A) sunt poziționate la cel puțin 4 baze față de semnalul poli(A). Acest fapt este în contrast cu cel al elementelor L1, la care semnalul poli(A) canonic AATAAA este imediat urmat de coada poli(A) (Belancio *et al.*, 2007). Astfel, semnalele non-canonice ar putea reprezenta substraturi pentru factori de poliadenilare diferiți decât cei folosiți de semnalele poli(A) canonice, iar acești factori ar putea dirija poliadenilarea printr-un mecanism diferit decât în cazul semnalelor poli(A) canonice.

Tabel 4. Semnale poli(A) potențiale folosite pentru poliadenilarea transcriptelor SVA 3' trunchiate

Grup	Semnale poli(A) potențiale			
	SVA A	SVA B	SVA C	SVA D
2			AATACA GATAGA (1)	AATAGA GATTGA (1)
3	TGTAGA AGTAGA (3)	TGTAGA AGTAGA (1)		
4			AAGAAA (2)	AAGAAA (5)
5	TCTACA (1)			
6				GTTAAA (1)
				GTTGAA (2)

Semnalele poli(A) potențiale sunt indicate în funcție de afilierea la subfamilie a fiecărui element

Numerele din paranteze indică numărul de elemente SVA care folosesc respectivul semnal de poliadenilare

NNNNNN – semnale poli(A) potențiale care diferă printr-o singură bază față de secvența de consens a subfamiliei

NNNNNN – semnale poli(A) potențiale care diferă de NNUANA, dar corespund cu secvența de consens a subfamiliei

NNNNNN – semnale poli(A) potențiale care diferă printr-o singură bază față de NNUANA și secvența de consens a subfamiliei

Lungimile cozilor poli(A) la transcriptele SVA 3' trunchiate exprimate în testiculul uman variază între 13 și 126 perechi de baze (Tabelul 5). Această variație poate fi atribuită amorării interne în etapa de revers-transcriere. Nu s-a putut stabili o corelație între lungimea cozii poli(A) și lungimea regiunii SINE-R. De asemenea,

lungimea cozii poli(A) pare să fie independentă de tipul de semnal poli(A) sau de afilierea la subfamilie a elementului care a generat transcriptul.

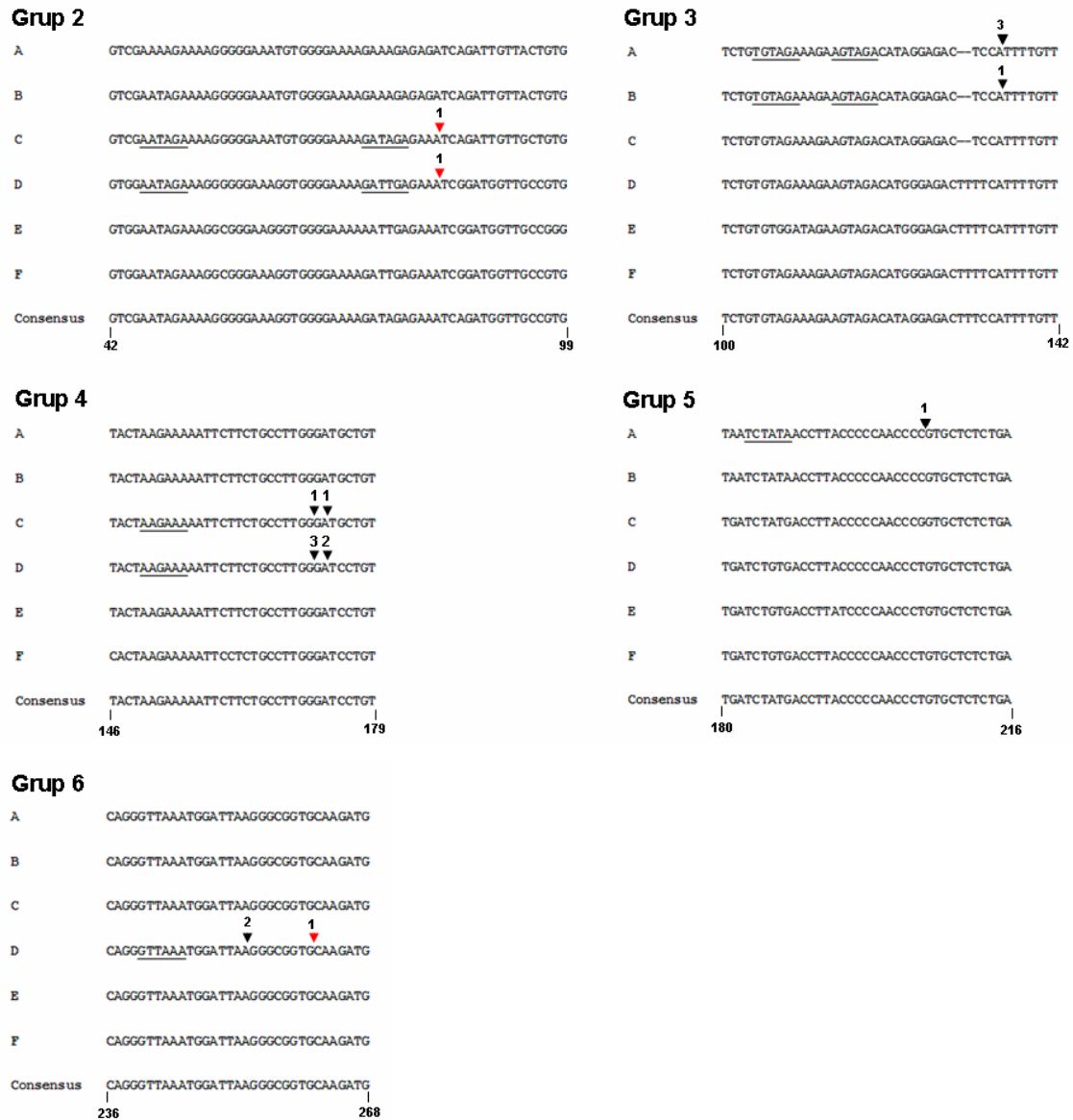


Figura 19. Situsuri alternative de clivare și poliadenilare dirijate de semnale poli(A) la transcriptele SVA exprimate în testiculul uman. Semnalele poli(A) (subliniate) și situsurile de clivare și poliadenilare (capetele de săgeți) sunt indicate în concordanță cu grupele de trunchiere ale elementelor SVA din genomul uman (vezi figura 14). Capetele de săgeți roșii indică situsurile de clivare și poliadenilare care diferă de cele observate la grupele de trunchiere din genomul uman. Numerele indică numărul de transcripte SVA care utilizează respectivul situs de clivare și poliadenilare. Limitele secvenței sunt indicate în secvența consens.

La patru dintre transcripte, lungimile cozilor poli(A) variază între 81 și 126 perechi de baze, ceea ce depășește lungimea cozilor poli(A) observate la elementele SVA 3' trunchiate din genom. La acestea din urmă, cea mai lungă coadă poli(A) este de 66 perechi de baze. Este probabil că aceste elemente SVA genomice au avut cozi poli(A) mai lungi la momentul inserării în genom care au suferit scurtări pe parcursul timpului (Chen *et al.*, 2005 și referințele cuprinse).

Table 5. Lungimea cozilor poli(A) la transcriptele SVA 3' trunchiate

Numele secvenței	Lungimea SINE-R (pb)	Semnalul poli(A) potențial	Lungimea cozii poli(A) (pb)	Elementul sursă	Afilierea la subfamilie
IB2L1	73	AATAGA/GATTGA	123	H19_70	SVA_D
IB3L10	83	AATACA/GATAGA	100	H2_418	SVA_C
IB3L4	133	TGTAGA/AGTAGA	81	H2_434	SVA_B
IB3L5	149	TGTAGA/AGTAGA	13	H12_A190	SVA_A
IB3L7	149	TGTAGA/AGTAGA	18		
IB3L9	149	TGTAGA/AGTAGA	14		
IB2L10	171	AAGAAA	126	H19_17	SVA_D
IB3L2	170	AAGAAA	16	nedeterminat	nedeterminat
IB3L6	171	AAGAAA	17	H2_456	SVA_D
IB4L1	171	AAGAAA	15	H19_105	SVA_C
IB4L5	171	AAGAAA	16	H14_A405	SVA_D
IB4L7	171	AAGAAA	25	H16_A573	SVA_D
IB3L8	173	AAGAAA	13	nedeterminat	nedeterminat
IB3L3	220	TCTACA	16	H7_1026	SVA_A
IB4L10	261	GTTAAA	16	H19_70	SVA_D
IB2L9	253	GTTGAA	14	H11_A52	SVA_D
IB3L1	253	GTTGAA	14		

Concluzii

1. Se presupune că eficiența mobilizării elementelor SVA este sporită de elemente Alu transduse (Damert *et al.*, 2009). Această ipoteză însă trebuie validată experimental. În această teză, ea a fost abordată parțial prin construcția a două seturi de vectori pentru testarea mobilizării SVA în culturi celulare. Un set de vectori conține un element SVA cu sau fără un element 3' Alu transdus pe lanțul plus. Al doilea set de vectori conține un element SVA cu sau fără un element 5' Alu transdus pe lanțul minus. Testarea acestor vectori în culturi celulare poate oferi dovada experimentală că elementele Alu transduse sporesc eficiența mobilizării elementelor SVA și de asemenea, ar putea elucida care orientare a elementului Alu este mai importantă pentru mobilizarea elementelor SVA mediată de L1.

2. A fost realizată o listă completă de elemente SVA 3' trunchiate adnotate din secvența genomului uman de referință. Majoritatea acestor elemente au cozi poli(A), indicând faptul că poliadenilarea prematură ar putea fi mecanismul responsabil pentru fenomenele de trunchiere la capătul 3'. Analiza elementelor a indicat că poliadenilarea apare cu precădere la anumite situsuri poli(A), iar secvențe potențatoare ar putea fi responsabile pentru această preferință.

Analiza genomului uman pentru elemente SVA 3' trunchiate a evidențiat noi variante structurale caracterizate prin deleții interne și prin inserții de retrotranspozoni în regiunea SINE-R. Acestea din urmă indică faptul că elementele SVA conțin secvențe țintă pentru endonucleaza elementului L1 în SINE-R.

Identificarea de transcripte SVA 3' trunchiate și poliadenilate confirmă ipoteza că poliadenilarea prematură este mecanismul responsabil pentru fenomenele de trunchiere la capătul 3' observate la elementele genomice SVA care poartă cozi poli(A). Situsurile poli(A) utilizate de transcriptele SVA 3' trunchiate exprimate în testiculul uman se aseamănă cu cele utilizate de elementele SVA 3' trunchiate din genom.

În final, s-a obținut dovada experimentală că subfamilile mai vechi SVA_A, B și C sunt transcrise în testiculul uman, indicând faptul că aceste subfamilii încă sunt capabile să ne modeleze genomul.

Bibliografie

- Beaudoing, E., S. Freier, J.R. Wyatt, J.M. Claverie, and D. Gautheret. 2000. Patterns of variant polyadenylation signal usage in human genes. *Genome Research* **10**: 1001-1010.
- Belancio, V.P. and P. Deininger. 2003. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nature Genetics* **35**: 363-366.
- Belancio, V.P., D.J. Hedges, and P. Deininger. 2008. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Research* **18**: 343-358.
- Belancio, V.P., M. Whelton, and P. Deininger. 2007. Requirements for polyadenylation at the 3' end of LINE-1 elements. *Gene* **390**: 98-107.
- Bennett, E.A., H. Keller, R.E. Mills, S. Schmidt, J.V. Moran, O. Weichenrieder, and S.E. Devine. 2008. Active Alu retrotransposons in the human genome. *Genome Research* **18**: 1875-1883.
- Borson, N.D., W.L. Salo, and L.R. Drewes. 1992. A lock-docking oligo(dT) primer for 5' and 3' RACE PCR. *PCR Methods and Applications* **2**: 144-148.
- Chen, J.-M., P.D. Stenson, D.N. Cooper, and C. Ferec. 2005. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. *Human Genetics* **117**: 411-427.
- Cordaux, R., and M.A. Batzer. 2009. The impact of retrotransposons on human genome evolution. *Nature Reviews. Genetics* **10**: 691-703
- Damert, A., J. Raiz, A.V. Horn, J. Lower, H. Wang, J. Xing, M.A. Batzer, R. Lower, and G.G. Schumann. 2009. 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome Research* **19**: 1992-2008.
- Danckwardt, S., I. Kaufmann, M. Gentzel, K.U. Foerstner, A.-S. Gantzert, N.H. Gehring, G. Neu-Yilik, P. Bork, W. Keller, M. Wilm et al. 2007. Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals. *The EMBO Journal* **26**: 2658-2669.
- Dewannieux, M., C.C. Esnault, and T. Heidmann. 2003. LINE-mediated retrotransposition of marked Alu sequences. *Nature Genetics* **35**: 41-48.

- Dewannieux, M. and T. Heidmann. 2005. Role of poly(A) tail length in Alu retrotransposition. *Genomics* **86**: 378-381.
- Esnault, C., J. Maestre, and T. Heidmann. 2000. Human LINE retrotransposons generate processed pseudogenes. *Nature Genetics* **24**: 363-367.
- Gogvadze, E.V., and A.A. Buzdin. 2005. New mechanism of retrogene formation in mammalian genomes: in vivo recombination during RNA reverse transcription. *Molekuliarnaia Biologiia* **39**: 364-373.
- Hancks, D.C., A.D. Ewing, J.E. Chen, K. Tokunaga, and H.H. Kazazian, Jr. 2009. Exon-trapping mediated by the human retrotransposon SVA. *Genome Research* **19**: 1983-1991.
- Hancks, D.C. and H.H. Kazazian, Jr. 2010. SVA retrotransposons: Evolution and genetic instability. *Seminars in Cancer Biology* **20**: 234-245.
- Hancks, D.C, J.L. Goodier, P.K Mandal, L.E. Cheung, and H.H. Kazazian, Jr. 2011. Retrotransposition of marked SVA elements by human L1s in cultured cells. *Human Molecular Genetics* **20**: 3386-3400
- Kim, H.S., B.H. Hyun, J.Y. Choi, and T.J. Crow. 2000. Phylogenetic analysis of a retroposon family as represented on the human X chromosome. *Genes & Genetic Systems* **75**: 197-202.
- Millevoi, S. and S.p. Vagner. 2010. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Research* **38**: 2757-2774.
- Mills, R.E., E.A. Bennett, R.C. Iskow, and S.E. Devine. 2007. Which transposable elements are active in the human genome? *Trends in Genetics: TIG* **23**: 183-191.
- Ono, M., M. Kawakami, and T. Takezawa. 1987. A novel human nonviral retroposon derived from an endogenous retrovirus. *Nucleic Acids Research* **15**: 8725-8737.
- Ostertag, E.M., J.L. Goodier, Y. Zhang, and H.H. Kazazian, Jr. 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. *American Journal of Human Genetics* **73**: 1444-1451.
- Raiz J., A. Damert, S. Chira, U. Held, S. Klawitter, M. Hamdorf, J. Löwer, W.H. Strätling, R. Löwer, G.G Schumann. In press. The non-autonomous retrotransposon SVA is *trans*-mobilized by the human LINE-1 protein

- machinery. *Nucleic Acid Research*.
- Roy-Engel, A.M., A.-H. Salem, O.O. Oyeniran, L. Deininger, D.J. Hedges, G.E. Kilroy, M.A. Batzer, and P.L. Deininger. 2002. Active Alu element "A-tails": size does matter. *Genome Research* **12**: 1333-1344.
- Shen, L., L.C. Wu, S. Sanlioglu, R. Chen, A.R. Mendoza, A.W. Dangel, M.C. Carroll, W.B. Zipf, and C.Y. Yu. 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**: 8466-8476.
- Venkataraman, K., K.M. Brown, and G.M. Gilmartin. 2005. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes & Development* **19**: 1315-1327.
- Wang, H., J. Xing, D. Grover, D.J. Hedges, K. Han, J.A. Walker, and M.A. Batzer. 2005. SVA elements: a hominid-specific retroposon family. *Journal of Molecular Biology* **354**: 994-1007.
- Xing, J., H. Wang, V.P. Belancio, R. Cordaux, P.L. Deininger, and M.A. Batzer. 2006. Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 17608-17613.