

Svoluji k zapůjčení své diplomové práce ke studijním účelům a prosím, aby byla vedena přesná evidence vypůjčovatelů. Převzaté údaje je vypůjčovatel povinen řádně ocitovat.

Univerzita Karlova v Praze  
Přírodovědecká fakulta  
Katedra genetiky a mikrobiologie

DIPLOMOVÁ PRÁCE  
**Vyhledávání genomických míst s trvalou expresí  
integrovaných retrovirů: přístup pomocí kuřecích nádorů  
indukovaných RSV**

Jan Kotáb

Praha 2009

**Školitel: RNDr. Jiří Hejnar, CSc.**

**Oddělení buněčné a virové genetiky, ÚMG AV ČR, v.v.i.**

**Prohlašuji, že tato diplomová práce byla vypracována samostatně pod dohledem školitele a že všechny použité prameny byly rádně ocitovány.**

.....

**Poděkování:**

Na tomto místě chci poděkovat svému školiteli RNDr. Jiřímu Hejnarovi, CSc. za trpělivost, se kterou vedl moji práci, Ing. Věre Hoserové a RNDr. Markétě Reinišové za obětavou pomoc při laboratorní činnosti, Ing. Jiřímu Plachému, CSc., Mgr. Kateřině Trejbalové Ph.D., Mgr. Janě Blažkové Ph.D., Mgr. Magdě Matouškové, Mgr. Petru Divinovi, Ph.D, a Mgr. Volodymyru Stepanetsovi Ph.D. za pomoc a informace a všem zaměstnancům Oddělení molekulární virologie za ochotu poradit a pomoci. V neposlední řadě děkuji za přátelské prostředí, ve kterém tato práce vznikala.

Tato diplomová práce vznikla v letech 2005-2009 v rámci grantového projektu  
Grantové agentury České republiky č. 204/07/1030.

## **Genome-wide screening of sequences permitting long-term expression of integrated retroviruses: a new approach using RSV-induced tumors in chicken.**

Retroviral integration has been shown to be non-random at the whole genome scale and integration preferences of many retroviruses are already known. The master factor of biased retroviral integration and target site selection is viral integrase, which is genus-specific. The chromosomal features directing integration site selection are more complicated and multiple factors play a role here. A better insight into the processes of biased retrovirus integration, particularly the genomic features of target sites permitting the long-term and efficient expression of integrated retrovirus and retrovirus-transduced genes, are important for application of retrovirus-derived vectors in transgenesis and gene therapy.

We have cloned and sequenced integration sites of Rous sarcoma virus (RSV) obtained from RSV-induced tumors in chickens and mapped them in the chicken genome assembly. Our genome-wide analysis displays much stronger preference for transcription units in comparison with our previous analysis of not-selected RSV integration sites. Our data set will be used for more detailed analysis of integration considering the proximity to transcription start sites, GC content, gene density and orientation of integrated retroviruses.

**Key words:** retrovirus, integration, target sites, transcriptionally active provirus

**Klíčová slova:** retrovirus, integrace, cílová místa, transkripčně aktivní provirus

## OBSAH:

<b>1. SEZNAM ZKRATEK .....</b>	<b>- 8 -</b>
<b>2. ÚVOD .....</b>	<b>- 9 -</b>
<b>3. PŘEHLED LITERATURY.....</b>	<b>- 11 -</b>
<b>3.1 RETROVIRY OBECNĚ .....</b>	<b>- 11 -</b>
<b>3.2 ŽIVOTNÍ CYKLUS.....</b>	<b>- 13 -</b>
<b>3.3. VIRUS ROUsoVA SARKOMU .....</b>	<b>- 16 -</b>
<b>3.4. INTEGRACE DO HOSTITELSKÉ DNA .....</b>	<b>- 17 -</b>
<b>3.4.1. Integrační mechanismus .....</b>	<b>- 17 -</b>
<b>3.5. BUNĚČNÉ FAKTORY OVLIVŇUJÍCÍ MÍSTO INTEGRACE.....</b>	<b>- 19 -</b>
<b>3.5.1. Primární struktura DNA .....</b>	<b>- 20 -</b>
<b>3.5.2. Přístupnost člověké DNA .....</b>	<b>- 21 -</b>
<b>3.5.3. Transkripční aktivita člověké DNA .....</b>	<b>- 22 -</b>
<b>3.5.4. Buněčný cyklus.....</b>	<b>- 23 -</b>
<b>3.5.5. Buněčné proteiny schopné vázat PIC .....</b>	<b>- 24 -</b>
<b>3.6. INTEGRAČNÍ PREFERENCE U RŮZNÝCH SKUPIN RETROVIRŮ .....</b>	<b>- 26 -</b>
<b>3.7. RETROVIROVÉ VEKTORY .....</b>	<b>- 29 -</b>
<b>4. MATERIÁL A METODY .....</b>	<b>- 30 -</b>
<b>4.1. PRÁCE S BUNĚČNÝMI KULTURAMI A VIRY .....</b>	<b>- 30 -</b>
<b>4.1.1. Použité buněčné linie .....</b>	<b>- 30 -</b>
<b>4.1.2. Viry.....</b>	<b>- 30 -</b>
<b>4.1.3. Používaná média .....</b>	<b>- 31 -</b>
<b>4.1.4. Kultivace buněk, rozmražování buněk.....</b>	<b>- 31 -</b>
<b>4.1.5. Sklizení buněk .....</b>	<b>- 31 -</b>
<b>4.1.6. Transfekce buněk .....</b>	<b>- 32 -</b>
<b>4.1.7. Sběr a zamražení viru.....</b>	<b>- 32 -</b>
<b>4.1.8. Tlitrace viru .....</b>	<b>- 32 -</b>
<b>4.1.9. Inokulace kuřat .....</b>	<b>- 33 -</b>
<b>4.2. MATERIÁL A METODY, PRÁCE S NUKLEOVÝMI KYSELINAMI .....</b>	<b>- 33 -</b>
<b>4.2.1 Použité vektory.....</b>	<b>- 33 -</b>
<b>4.2.2. Izolace chromozomální DNA .....</b>	<b>- 34 -</b>
<b>4.2.3. I-PCR .....</b>	<b>- 34 -</b>
<b>4.2.3.1. Princip I-PCR .....</b>	<b>- 34 -</b>
<b>4.2.3.2. Postup .....</b>	<b>- 36 -</b>
<b>4.2.4. Izolace PCR produktu z gelu.....</b>	<b>- 37 -</b>
<b>4.2.5 Ligace PCR produktu s místem integrace do vektoru pGEM-T Easy .....</b>	<b>- 38 -</b>
<b>4.2.6. Příprava kompetentních bakterií .....</b>	<b>- 38 -</b>
<b>4.2.7. Transformace bakterií.....</b>	<b>- 39 -</b>
<b>4.2.8. Koloniové PCR .....</b>	<b>- 39 -</b>
<b>4.2.9. Sekvenace DNA .....</b>	<b>- 40 -</b>
<b>4.2.10. Minipreparace DNA .....</b>	<b>- 42 -</b>
<b>4.2.11. Použitý počítačový software .....</b>	<b>- 42 -</b>
<b>4.2.12. Mapy použitých vektorů .....</b>	<b>- 43 -</b>

<b>5. VÝSLEDKY .....</b>	<b>- 45 -</b>
<b>5.1. EXTRAKCE A OVĚŘENÍ MÍST INTEGRACE .....</b>	<b>- 45 -</b>
<b>5.2 DISTRIBUCE INTEGRAČNÍCH MÍST NA CHROMOZOMECH .....</b>	<b>- 46 -</b>
<b>5.3. PREFERENCE RSV PRO GENY .....</b>	<b>- 59 -</b>
<b>6. DISKUSE.....</b>	<b>- 60 -</b>
<b>7. SHRNUTÍ .....</b>	<b>- 62 -</b>
<b>8. SEZNAM LITERATURY .....</b>	<b>- 63 -</b>
<b>9. PŘÍLOHA.....</b>	<b>- 68 -</b>

## **1. SEZNAM ZKRATEK**

**ASLV** Avian sarkoma leukosis virus  
**AV ČR** Akademie věd České republiky  
**BAF** Barrier-to-autointegration factor  
**bp** pár bází  
**cDNA** komplementární deoxyribonukleová kyselina  
**CEF** kuřecí embryonální fibroblasty  
**DNA** deoxyribonukleová kyselina  
**FV** Foamy virus  
**HERV** Human endogenous retrovirus  
**HIV** Human immunodeficiency virus  
**HSC** (hematopoietic stem cell)  
**HTLV** Human T-cell leukemia  
**Kb** kilobáze  
**LTR** long terminal repeat  
**Mb** megabáze  
**MLV** Murine leukemia virus  
**MMTV** (Mouse mammary tumor virus)  
**obr.** obrázek  
**PCR** polymerázová řetězová reakce  
**PERV** Porcine endogenous retrovirus  
**PIC** Preintegrační komplex  
**RNA** ribonukleová kyselina  
**RSV** virus Rousova sarkomu  
**RT** (Room Temperature) pokojová teplota  
**SIV** Simian immunodeficiency virus  
**tab.** tabulka  
**tRNA** transferová ribonukleová kyselina  
**X-SCID** X-linked severe combined immunodeficiency

## 2. ÚVOD

Virus Rousova sarkomu (RSV) je akutně transformující retrovirus patřící do komplexu ptačích sarkomových a leukosových virů (ASLV) charakteristických přítomností virového onkogenu *v-src*. RSV, stejně jako ostatní retroviry, potřebuje pro úspěšnou replikaci vložit kopii svého genomu ve formě DNA do genomu cílové buňky.

Jak se ukázalo v posledních letech, není tento proces zcela náhodný. Celkově se dá říci, že integrace u retrovirů není sekvenčně specifická, ale probíhá s určitými preferencemi na úrovni sekundární struktury DNA a chromatinové struktury.

Retroviry po vstupu do hostitelské buňky, uvolnění obsahu kapsidy a reversní transkripcí virové genomové RNA interagují s buněčnými proteiny a tvoří preintegrační komplex (PIC), který se pak stává hlavním integračním činitelem. Interakce PIC a genomové DNA zprostředkuje zejména integráza, která se mezi jednotlivými retroviry liší, takže můžeme pozorovat rozdílné integrační preference jednotlivých retrovirových rodů. Integráza ale není jediným činitelem a na konečném umístění provirové DNA se podílí též řada buněčných faktorů.

Důležitost porozumění procesu integrace se naplno projevila při tvorbě vektorů pro genovou léčbu a expresních vektorů, kdy jsou na retrovirové vektory kladený specifické požadavky směřující k eliminaci jejich genotoxicity a zároveň zajištění jejich dlouhodobé exprese.

Integrační preference většiny retrovirů v buněčných kulturách jsou již charakterizovány. O RSV je známo, že stejně jako ostatní ASLV mírně preferuje oblasti bohaté na geny a oblasti s vysokým zastoupením GC. Pro genovou terapii je podstatné, že RSV a vektory odvozené z ASLV nemají významnou afinitu pro počátky transkripce jako je tomu zejména v případě MLV. V případě retrovirové transdukce genů pro expresi rekombinančních proteinů nebo pro genovou terapii se však nejedná o pouhou integraci, ale též o selekci buněk, které v dostatečné míře a po dostatečně dlouhou dobu transdukované geny exprimují. Je tedy třeba studovat též retrovirovou integraci v selektovaných buněčných populacích a popsat integrační místa, ve kterých je takto produktivní integrace retrovíru umožněna.

Naše práce se zaměřila na charakterizaci integračních míst RSV z povětšinou monoklonálních nádorů sarkomového typu indukovaných v přirozeném hostiteli

tohoto viru, kterým je kur domácí (*Gallus gallus*). Jelikož již byly v naší laboratoři studovány integrační preference RSV v infikovaných a neselektovaných kuřecích embryonálních fibroblastech (CEF), nabízí se možnost porovnání integračních míst v buňkách podrobených selekci *v in vivo* rostoucích nádorech. Především nás zajímalo, jestli dojde k posunu integrace do aktivních genů a stabilně transkribovaných oblastí, což by mělo jít na vrub nádorové selekci, kdy jen integrace do vhodného místa genomu zajistí transkripci genu *v-src* na úrovni potřebné pro vznik sarkomu. S tím by měl souviseť také větší počet integrací do genů, u nichž byla prokázána zvýšená hladina exprese v pojivových tkáních.

Pro tuto práci bylo třeba zkonstruovat pantropní retrovírový vektor, který nám umožnil infikovat kuřata linie P bez endogenních retrovirů. K tomuto kroku jsme museli přistoupit po počátečních neúspěších, které byly zapříčiněny defektním receptorem Tvc, který způsobuje semirezistenci linie P pro virovou podskupinu C..

Dalším cílem bylo získat inokulaci pseudotypovaným RSV vektorem co nejvíce virem indukovaných monoklonálních nádorů, izolovat z nich chromozomální DNA a pomocí I-PCR techniky identifikovat genomická místa integrace. Takto získaná sada integračních míst bude materiélem pro pozdější genomickou analýzu.

Výsledky této práce by měly pomoci porozumět interakcím integrovaného RSV a genomické DNA hostitele a celkově přispět ke snaze optimalizovat retrovírové vektory pro genovou terapii nebo pro expresi rekombinantních proteinů.

### **3. PŘEHLED LITERATURY**

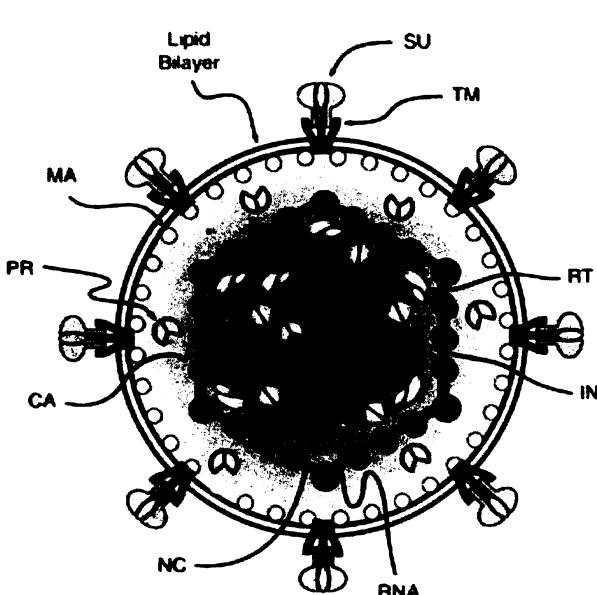
#### **3.1 Retroviry obecně**

Historie retrovirologie se píše již skoro sto let a je úzce spjata s výzkumem akutně onkogenních retrovirů. Roku 1911 Peyton Rous publikoval, že neznámá filtrabilní agens způsobuje sarkomy u kuřat a že za těmito nádory stojí virus, který se dnes jmenuje po svém objeviteli - virus Rousova sarkomu (RSV). Ten se od té doby stal předmětem mnoha studií, které pomohly formulovat odpovědi na celou řadu otázek týkajících se retrovirů, jejich odlišnosti od ostatních skupin virů a jejich roli v onkogenezi. Dnes je již známo několik desítek druhů retrovirů a nové druhy a kmeny se stále objevují.

Retroviry se od ostatních virů liší tím, že jsou funkčně diploidní a namísto jedné tak mají hned dvě nekovalentně spojené kopie +RNA s methylguanosinovou čepičkou a polyA sekvencí a zejména pak svým životním cyklem, ve kterém dochází nejprve k reverzní transkripci a následně k integraci cDNA verze virového genomu zvané provirus do DNA. Typická retrovirová partikule je složena z obalu, skrze který vyčnívají do prostoru povrchové glykoproteiny ukotvené pomocí transmembránových glykoproteinů, matrix proteinu, kapsidy, dvou molekul RNA, které jsou asociovány s nukleokapsidovým proteinem, jenž pomáhá s dimerizací obou molekul a s navázáním tRNA primeru a z několika desítek molekul integrázy, reverzní transkriptázy a proteázy (obr. 1).

Virová RNA je přibližně 10 kb dlouhá, na 5' konci opatřena methylguanosinovou čepičkou, po níž následuje terminální repetice R a unikátní 5' sekvence U5. K té přiléhá primer vázající místo (primer binding site - PBS), kde dochází při reverzní transkripcí k nasednutí specifické tRNA, která následně poslouží jako primer pro minus DNA řetězec proviru. Následují kódující sekvence virových proteinů - group specific antigen (*gag*), proteáza (*pro*) a polymeráza (*pol*), které kódují proteiny virové kapsidy a sekvence envelope (*env*), která kóduje proteiny virového obalu. Celkem je pak z těchto sekvencí vytvořeno minimálně 8 virových proteinů - matrixový protein (MA), kapsidový protein (CA), nukleokapsidový protein (NC), proteáza (PR), reverzní transkriptáza (RT), integráza (IN), povrchový glykoprotein (SU) a transmembránový glykoprotein (TM). Sekvence mezi 5' LTR a

*gag* obsahuje enkapsidační signál, který obstarává sbalení virové RNA do kapsidy a oblast *gag-pol* slouží také jako fakultativní intron. Za sekvencí *env* následuje polypurinový trakt (PPT), který funguje jako primer pro plus řetězec DNA. Proximální část je pak zakončena unikátní sekvencí U3 a 3' terminální repeticí R, za kterou RNA končí polyA repeticí.



Obrázek 1. Schematický řez retrovirovou partikulí. Transmembránový (TM) a povrchový (SU) glykoprotein spojený disulfidickou vazbou na lipidové dvouvrstvě, matrixový (MA) protein, kapsidový (CA) protein, nukleokapsidový (NC) protein a proteáza (PR). Uvnitř kapsidy je reverzní transkriptáza (RT), integráza (IN) a 2 molekuly RNA.

Převzato od Coffin et al. 1997

Retroviry samotné pak dělíme na exogenní a endogenní. Endogenní retroviry vznikly nejspíše integrací exogenního retrovíru do germinálních buněk hostitele. Naprostá většina z nich se pak časem stala defektními. Patří mezi ně například HERVs (Human endogenous retroviruses), které tvoří asi 8% lidské DNA. Některé z nich se dokonce staly endosymbionty a u placentálů je endogenní retrovirus zodpovědný za fúzi buněk placenty (MI et al. 2000).

Retrovirům příbuzné jsou retrotransposony, což jsou mobilní elementy eukaryot, které nekódují obalový protein a tudíž postrádají extracelulární formu ve svém životním cyklu. Dělí se na virové retrotransposony s LTR a na nevirové retrotransposony bez LTR. LTR retrotransposony kódují vlastní reverzní transkriptázu integrázu a *gag* protein. Dělí se na retrovirům příbuznější Ty3/gypsy (*Metaviridae*) a Ty1/copia (*Pseudoviridae*) (HAVECKER et al. 2004).

Mezi retroelementy bez LTR patří v lidském genomu hojně zastoupené a autonomní LINE (long interspersed elements) a neautonomní SINE (short interspersed elements).

Retrotransposony kvasinek posloužily jako výborný model pro studium mechanismů cílcích integraci a zjištěné integrační preference pro oblasti heterochromatinu, telomer a promotorové oblasti genů přepisovaných DNA polymerázou III a II dokreslují důležitost místa integrace v kontextu s životní strategií konkrétního retroelementu (ZHU et al. (2003), BOEKE et al. (1998), BOWEN et al. (2003), BUSHMAN (2003)).

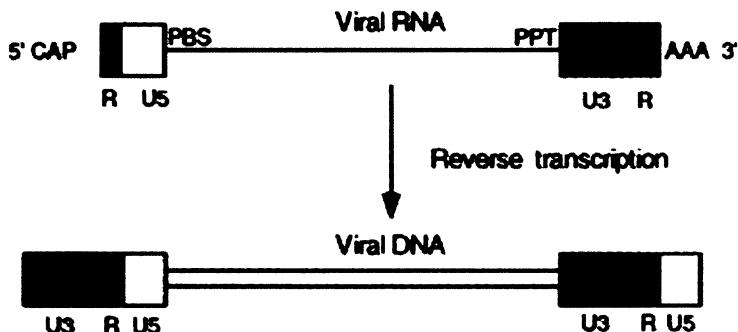
### 3.2 Životní cyklus

Životní cyklus u retrovirů (obr. 3) začíná navázáním virových povrchových proteinů na specifický receptor na povrchu plazmatické membrány hostitelské buňky. Tato interakce vede ke konformační změně virových glykoproteinů, která následně zprostředkuje fúzi lipidových dvouvrstev buněčné a virové membrány a umožní tím vstup virové partikule do cílové buňky. V případě endocytózy virové částice se virus uvolní změnou pH endozomu.

Následujícím krokem je reverzní transkripce (obr. 2), reakce, která se v době svého objevu přičila dogmatu o toku genetické informace ve směru DNA => RNA a která dala retrovirům jejich současný název. Reverzní transkripce probíhá v cytoplazmě hostitelské buňky, kde virová RNA tvoří nukleoproteinový komplex. Tento, zatím ne zcela charakterizovaný komplex, vytvoří za pomoci reverzní transkriptázy (což je virová RNA dependentní DNA polymeráza) a RNasy H lineární duplex DNA. Takto vytvořená virová dsDNA je přesnou kopíí původní RNA, ale navíc nese na obou koncích repetitivní sekvence zvané long terminal repeats (LTR). Ty vznikají během reverzní transkripce, kdy dochází postupně ke dvěma přeskokům transkriptního komplexu a tvorbě provirové dsDNA.

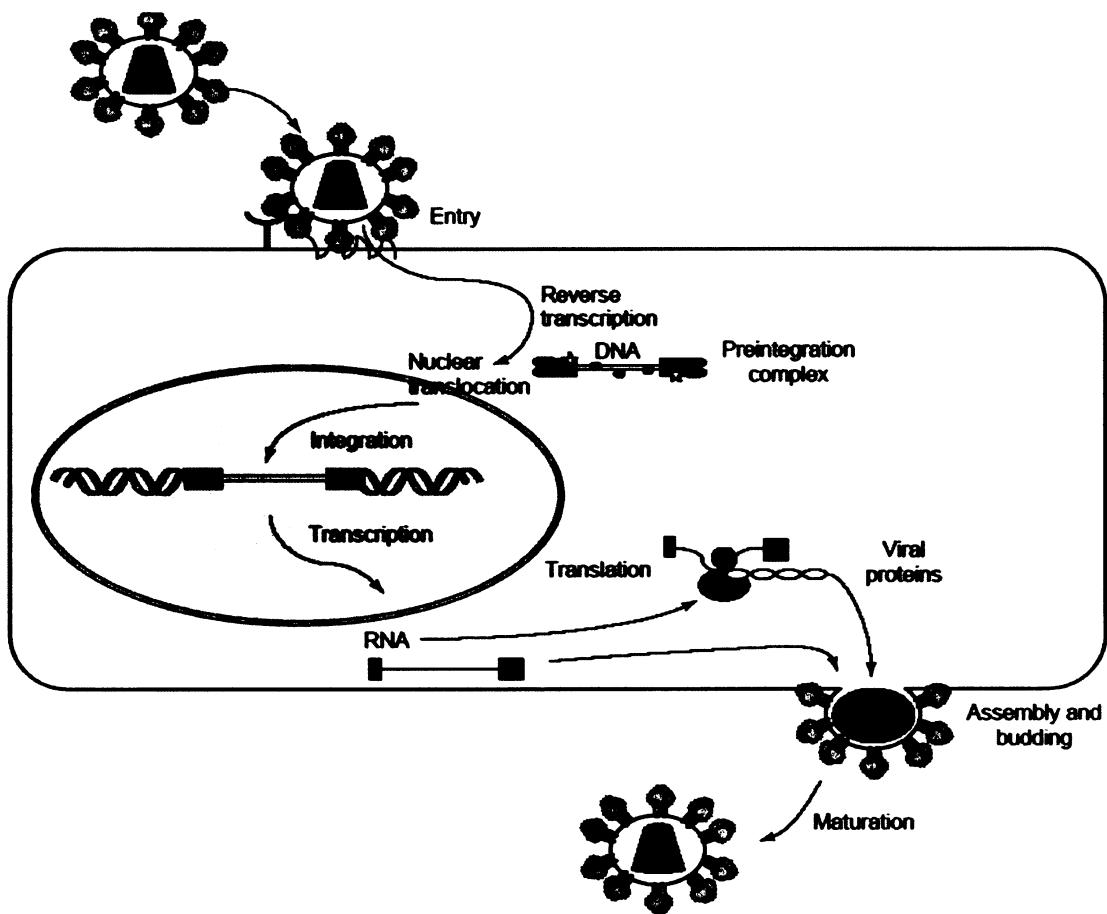
Dokončením reverzní transkripce je odstartována další část životního cyklu a tou je integrace virové dsDNA do genomu hostitelské buňky. Integrovaná forma viru se nazývá provirus a jedině úspěšná integrace zaručuje retrovíru transkripcí jeho genomu a tvorbu infekčních partikulí a v případě integrace do buněk germinální linie

i jeho přenos na následující generace. Tento krok je podrobněji popsán v kapitole 3.4., nebudu se mu proto zde více věnovat.



**Obrázek 2.** Reversní transkripcie virového RNA genomu vytvoří lineární DNA duplex. Je zde vyznačena pozice sekvencí R, U5, a U3, polypurinového traktu (PPT) a primer-vázající sekvence (PBS). Reverzní transkripcie duplikuje oblasti U5 a U3 takže DNA produkt je na obou koncích delší než původní RNA. Převzato od Coffin et al. (1997)

Po úspěšné integraci nastává reprodukční fáze, kdy dochází k transkripci virové DNA a tvorbě virových mRNA. Ta je řízena z 5' LTR, kde se nachází virový promotor a enhancer, který určuje tkáňový tropismus viru a často bývá přítomen ve dvou v tandem opakovaných kopiích. V 3' LTR se pak uplatňuje polyadenylační signál. Virus se v této fázi vývoje spoléhá téměř výlučně na transkripční aparát hostitelské buňky, na buněčnou polymerázu II a její transkripční faktory. Zároveň pomocí svého velmi silného promotoru mnohonásobně zvyšuje hladinu exprese virových mRNA. Některé retroviry pak kódují vlastní transkripční aktivátory, například proteiny Tat a Tax virů HIV a HTLV, jimž dále zvyšují hladinu exprese proviru. Ta začíná z 5'LTR, z které jsou přepsány oblasti R a U5 a po osazení 5' nascentní RNA metylguanosinovou čepičkou a přepsání kódujících sekvencí proviru včetně U3 a R 3'LTR, je transkript přestříhnut a 3'OH konec je polyadenylován.



**Obrázek 3.** Shrnutí životního cyklu retrovirů. Virion se naváže na buněčnou membránu za pomocí konkrétního receptoru (v případě zde popsánoho HIV interaguje se dvěma receptory: s CD4 molekulou a pomocnou molekulou CCR5, nebo CXCR4), následuje fúze membrán a uvolnění virového jádra do cytosolu hostitelské buňky. Pak je virová RNA přeložena reverzní transkripcí do lineární dsDNA virem kódovanou reverzní transkriptázou. Virová DNA interahuje s virovými a buněčnými proteiny za vzniku preintegračního komplexu (PIC) a migruje do jádra, kde ji virová integráza integruje do genomu hostitele. Buněčná RNA polymeráza II pak nasynthesizuje virové RNA. Následuje syntéza proteinů, složení virionů a vypuštění z membrány. Vše je zakončeno maturací virionů, kdy jsou virovou proteázou naštěpeny finální produkty.

Převzato od Ciuffi a Bushman (2006)

Primární transkript je poté podroben postranskripčním úpravám aparátem hostitelské buňky, kdy je kombinován alternativní sestřih s polyproteinovou strategií a se supresí terminačního tripletu, nebo posunem čtecího rámce. Kompletní transkript opatřený čepičkou na 5' konci a polyA na 3' konci pak slouží jako nová virová genomová RNA. Poměr mezi mRNA, která je sestřížena na subgenomovou mRNA a která putuje do cytoplazmy v nesestřížené podobě jako genomová RNA, je pak u složitějších retrovirů určován pomocí proteinů Rex a Rev. Obalové proteiny, které vznikají z obalového polyproteinu *env*, který byl přeložen ze subgenomové

RNA s odstříženým fakultativním intronem *gag-pol* jenž obsahuje i enkapsidační signál, nejsou enkapsidovány a interakcí s buněčnými receptory chrání buňku před virovou superinfekcí.

Po úspěšné transkripcí genomové RNA a translaci obalových glykoproteinů tak může nastat poslední fáze životního cyklu, kterou je morfogeneze retrovirové partikule. Ta může nastat buď v cytoplazmě, nebo na vnitřním povrchu cytoplazmatické membrány a nebo na povrchu vnitřních buněčných membrán. Nejprve dojde k sestavení "nezralých" nukleokapsid obsahujících polyproteiny Gag a Gag-Pol, dvě molekuly virové RNA a tRNA primery. Ty následně vypučí z plazmatické membrány v místech s připravenými virovými obalovými proteiny SU a TN. Proces je zakončen maturací nukleokapsidy spočívající v proteolytickém rozstříhání polyproteinu Gag a Gag-Pol na finální produkty (MA, CA, NP, PR, RT, IN) a v dimerizaci molekul virové RNA asociované s tRNA primerem.

### 3.3. Virus Rousova sarkomu

Virus Rousova sarkomu je členem rodu *Alpharetrovirus* a od ostatních virů z této skupiny se odlišuje tím, že kromě klasické trojice kódujících sekvencí – *gag*, *pol* a *env*, nese v genomu ještě gen *v-src*. *V-src* je původem buněčný gen kódující tyrozin- proteinkinazu, který se v minulosti stal součástí viru a zároveň přišel o svou C-terminální inhibiční doménu. Tím se z proto-onkogenu *c-src* stal akutně onkogenní *v-src*, který není pro životní cyklus RSV nezbytný, ale činí z něj akutně transformující virus. RSV je jediný transformující retrovirus, který rekombinací s onkogenem neztrácí replikační kompetenci. Spolu s ostatními ASLV viry sdílí *env* protein, který je zodpovědný za tvorbu transmembránového a povrchového glykoproteinu a díky jehož různorodosti dělíme tyto viry do podskupin A,B,C,D,E a J. Jednotlivé podskupiny sdílejí určitý okruh potenciálních hostitelů (BARNARD et al. 2006).

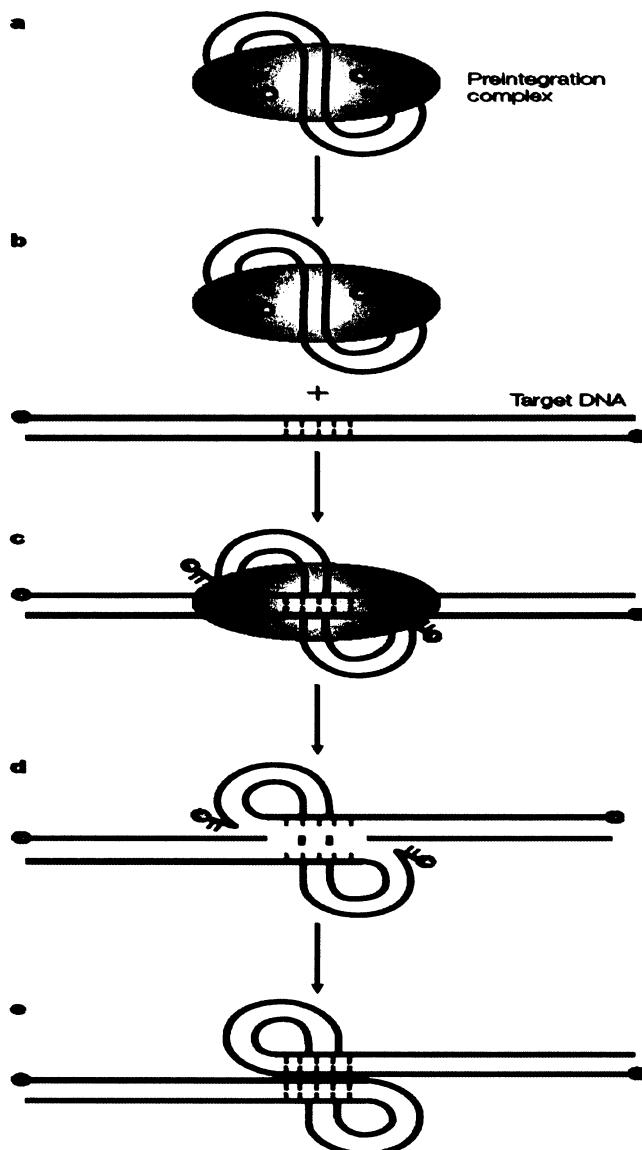
### **3.4. Integrace do hostitelské DNA**

Integrace je základním krokem v životním cyklu retrovíru, během kterého dojde, v jádru hostitelské buňky, ke vzniku kovalentní vazby mezi virovou cDNA a hostitelskou chromozomální DNA a je nezbytným krokem k ustanovení produktivní infekce.

#### **3.4.1. Integrační mechanismus**

Srdcem celého integračního mechanismu je virová integráza (člen D,D(35)E transposase/IN superrodiny) (SANDMEYER 2003), která je plně funkční ve formě tetrameru (FAURE et al. 2005). Ta se po úspěšné reverzní transkripcí stane spolu s virovou cDNA, virovou matrix a některými proteiny nukleokapsidy součástí preintegračního komplexu (PIC) (FARNET et al. 1991, MILLER et al. 1997). V cytoplazmě hostitelské buňky nejprve přeruší a následně zkrátí provirovou DNA o dva nukleotidy přiléhající ke konzervované sekvenci CA na 3'koncích a umožní tím budoucí kontakt LTR DNA 3'OH konců s hostitelskou DNA. Poté následuje import do jádra, kde dojde k přerušení fosfodiesterové vazby v hostitelské DNA a transesterifikační reakci, která spojí 3'LTR s naštěpenými 5' konci. Následuje ligace a oprava vzniklých mezer za pomoci buněčných opravných enzymů (KATZ, RA et al. 1990, Li et al. 2006) (obr. 4).

Virová integráza je kromě posledního kroku (oprava mezer v DNA) *in vitro* schopna uskutečnit integraci sama o sobě, ovšem v živých systémech kooperuje s dalšími proteiny a tvoří tak PIC. Pozitivní vliv takovýchto interakcí na efektivitu integrace *in vitro* pozoroval například CARTEAU et al. (1999). Dá se předpokládat, že jejich funkce *in vivo* bude obdobná a nedávný výzkum pak ukázal, že některé z proteinů mohou být jedním z hlavních faktorů ovlivňujících výběr místa integrace proviru (CIUFFI et al. 2005, BOTBOL et al. 2008).



Obrázek 4. Integrace virové cDNA do genomu hostitele:

Hnědý ovál je tetramer virové integrázy a ostatní proteiny tvořící PIC, červené čáry značí virovou DNA, modré pak cílovou DNA.  
a | Komplementární DNA (cDNA) molekula s tupými konci.

b | Ještě než začne integrace, odstraní integráza 2 krajní nukleotidy z obou 3' konců.

c | 3' konce obsahující hydroxylové skupiny jsou připojeny k vyčnívajícím 5' koncům cílové DNA, rozštěpeným transesterifikační reakcí.

d | Nepárování bází vytvoří mezeru mezi hostitelskou a virovou DNA na obou řetězích.

e | Nakonec proběhne oprava mezer za účasti buněčného opravného aparátu a vznikne kompletní zaintegrovaný provirus ohraničený z obou stran 4-6bp repetitive cílové DNA.

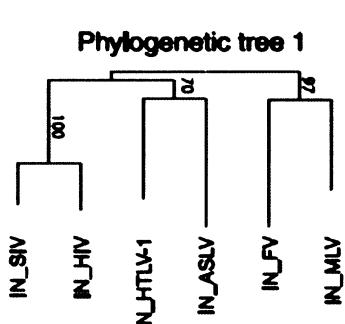
Převzato od Bushman et al. (2005)

Počet integrací do genomu hostitelské buňky se u jednotlivých skupin virů liší, neodvíjí se jen od počtu cDNA kopií viru v buňce. U MLV bylo pozorováno 9-14 integrací do jedné buňky (ODAWARA et al. 1998), u RSV 4-6 integrací do haploidního genomu (TSURUO a BALUDA 1977) ale u HIV-1 pouze 1-2 integrace (BUTLER et al. 2001). Touto výjimkou se zabývali ROSENBLUH et al. (2007) a LEVIN et al. (2009) a zjistili, že virový Rev protein se účastí regulace aktivity HIV-1 integrázy, která jeho vliv inhibuje pomocí dvou vlastních peptidů.

Integrace virové cDNA do genomu hostitele není sekvenčně specifická a retroviry se mohou integrovat prakticky na jakémkoli místě na chromozomu

(BUSHMAN et al. 2005). Výzkumy provedené v posledních letech na velkém množství integračních míst a s různými druhy retrovirů a mobilních elementů ale ukázaly, že retroviry se integrují s určitými charakteristickými preferencemi (ELLEDER et al. (2002), SCHROEDER, A. et al. (2002), FERBER et al. (2003), WU et al.(2003), HEMATTI et al.(2004), NAREZKINA et al. (2004))

Virová integráza se pak dá označit za hlavního druhově specifického činitele při výběru místa integrace. DERSE et al. (2007) vyšetřil sekvenční podobnost aminokyselin několika retrovirových integráz a vzniklý fylogenetický strom ukázal, že retroviry s příbuznými integrázami a doménami sdílejí i podobné integrační preference (obr. 5):



**Obrázek 5.** Fylogenetický strom vytvořený na základě srovnání aminokyselinové sekvence integráz.

Převzato od Derse et al. (2007)

### 3.5. Buněčné faktory ovlivňující místo integrace

Integrace virové cDNA do genomu hostitele není sekvenčně specifická a retroviry se mohou integrovat prakticky na jakémkoli místě na chromozomu (BUSHMAN et al. 2005). Výzkumy provedené v posledních letech na velkém množství integračních míst a s různými druhy retrovirů a mobilních elementů ale ukázaly, že retroviry se integrují s určitými unikátními preferencemi (ELLEDER et al. (2002), SCHROEDER et al. (2002) , FERBER et al. (2003), WU et al.(2003), HEMATTI et al. (2004), NAREZKINA et al. (2004)).

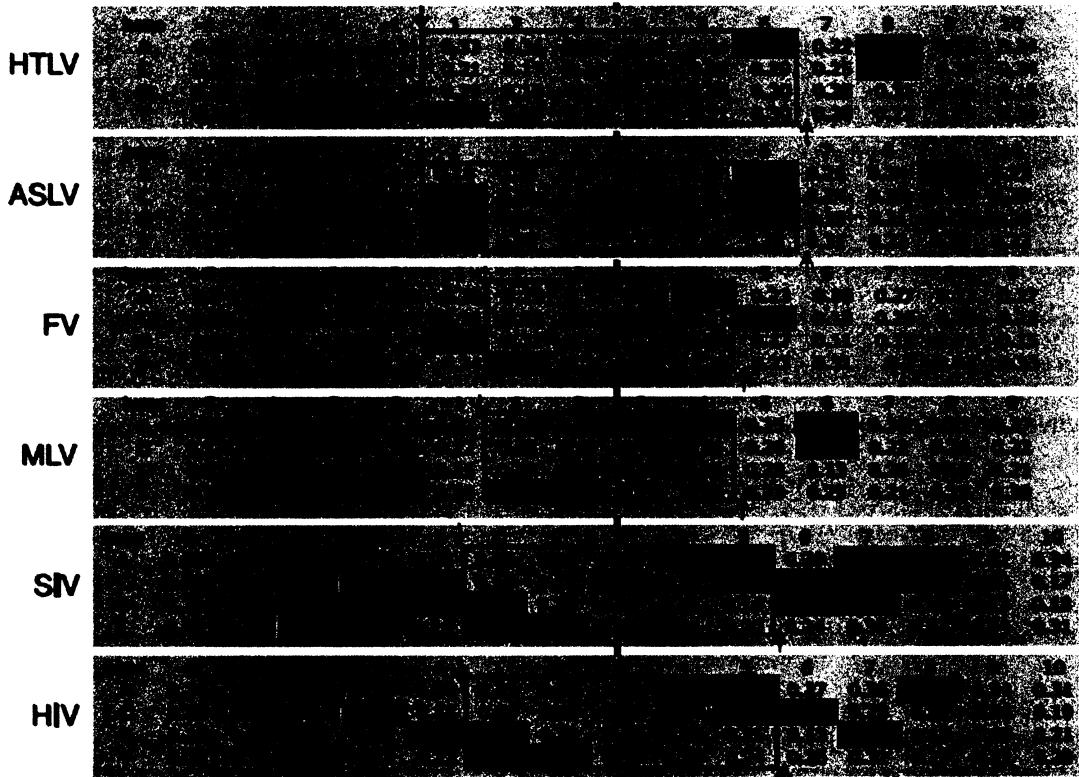
Postupně se tak zjistilo, že na místo integrace může mít vliv hned několik na sobě nezávislých faktorů a že integrační preference se u jednotlivých retrovirů, i ostatních mobilních elementů značně liší a to zejména v souladu s jejich životní strategií (WU a BURGESS 2004).

### **3.5.1. Primární struktura DNA**

Vliv primární struktury hostitelské DNA byl charakterizován již v prvních *in vitro* pracích (BOR et al. 1996). Postupně tak byly odhaleny slabé preference, které se však nedaly považovat za důležitého činitele v cílení integračního mechanismu. Že je s primární strukturou cílové DNA ale třeba také počítat, ukázal WU et al. (2005). Ve své práci porovnali sekvence *in vivo* integračních míst několika retrovirů (HIV-1, SIV, MLV, ASLV) a nalezli palindromické sekvence se středem v místě integrace, které měly charakteristickou konsensní sekvenci pro každý virus (obr. 6). Postupně byly tyto výsledky potvrzeny i u FV (NOWROUZI et al. 2006) a PERV (MOALIC et al. 2006). Takovýto palindrom se totiž jeví jako ideální substrát pro štěpení integrázou, která pracuje ve formě tetrameru (DEDELIS et al. 2007)

Vliv palindromické sekvence na výběr místa integrace pak ještě zdůraznil DERSE at al. (2007). Nejprve charakterizovali integrační preference HTLV-1 a poté porovnali palindromické vzorce jednotlivých druhů retrovirů mezi sebou a dokázali, že příbuzné retroviry, jenž mají tudíž i podobnou integrázou, se integrují do shodného palindromického vzorce (obr. 6).

Tato fakta ukazují na primární sekvenci jako na velmi důležitého činitele, zejména co se výběru místa integrace samotným retrovirem týče. Důležité je, aby cílová sekvence vyhovovala tetrameru konkrétní integrázy



**Obrázek 6.** Palindromické sekvence v místech retrovirové integrace: Integrace proběhla mezi pozicemi -1 a 1 na vrchním řetězci. Barevné obdélníky značí místa se statisticky odlišnou frekvencí bází od náhodně generovaných bází. ( $P < 0.01$ ), což je 30%/20%/20%/30% pro A/C/G/T. Báze s frekvencí větší o více jak 10% jsou zelené a báze u kterých došlo ke snížení frekvence o více, jak 10% jsou červené. Duplikovaná cílová sekvence je v modrém rámečku, a přenos DNA řetězce proběhl na pozici značené šípkami. Je zde vidět palindromický konsensus, tvořený preferencemi bází se středem ve zdvojené cílové sekvenci. Symetrie je vyznačena fialovou čárou.

Převzato od Derse et al. (2007)

### 3.5.2. Přístupnost cílové DNA

Přístupnost hostitelské DNA pro PIC je ovlivňována řadou faktorů. Jedním z nich je sekundární struktura DNA, kdy PRICIAK a VARMUS (1992) zjistili, že integrace nastává téměř výhradně ve velkém žlábku, a také ohnutí DNA, kdy je preferována DNA s větším zakřivením (PRUSS et al. 1994a). Dalším faktorem je pak sbalení cílové DNA do nukleozomů, kde je zřejmá preference pro DNA namotanou okolo nukleozomů v 10 nm (beads-on-a-string) konformaci, která napomáhá integraci ohnutím DNA (PRUSS et al. 1994b). Naopak sbalení do kompaktnějších forem integraci znesnadňuje (LEWINSKI et al. 2005). Totéž platí i o vlivu ohybu DNA indukovaném proteiny se schopností vázat DNA, který může

usnadnit integraci (MULLER a VARMUS 1994). Proteiny navázané na DNA mohou ale také svou přítomností zabránit nasednutí integračního komplexu (BOR et al. 1995). Většina těchto prací však byla provedena *in vitro* a nezahrnuje tak rozličné aspekty nitrobuněčného prostředí, které mohou mnohé z těchto faktorů dále modifikovat. Novější *in vivo* studie pak u HIV, MLV a slabě i u ASLV zjistily preferenci pro transkribované oblasti (HIV, ASLV) a pro transkripční starty (MLV) a tím i podpořily teorii o přístupnosti chromatinu (MITCHELL et al. 2004). Přístupnost cílové DNA a její deformace ohybem kolem nukleozomů, či DNA vázajících proteinů je tedy důležitým obecným faktorem, jenž je společný všem retrovirům, ale nevysvětuje rozdílné integrační preference *in vivo*.

### 3.5.3. Transkripční aktivita cílové DNA

Práce na toto téma provedená na retrovиру RSV ukázala, že vysoká hladina exprese, může působit jako inhibitor integrace (WEIDHAAS et al. 2000). Novější výzkum na tomtéž retrovirusu, pouze na jiné buněčné linii tento výsledek potvrdil, když po stonásobném zvýšení transkripční aktivity pozorovali šestinásobné snížení počtu integrací (MAXFIELD et al. 2005). Autoři těchto prací vysvětlují tento výsledek obsazením cílové DNA transkripční mašinérií a vznikem sterické překážky pro integrázu nebo tím, že v době transkripce není cílová DNA k dispozici ve dvouřetězcové formě.

Tyto výsledky jsou však v rozporu s jinými pracemi na RSV, které se zabývaly zejména charakterizací integračních míst RSV a ve kterých autoři zjistili, že RSV má slabou preferenci pro geny, ale bez zřejmého vlivu transkripční aktivity.(MITCHELL et al. 2004, NAREZKINA et al. 2004). Tento rozpor je vysvětlován buď tím, že za integračním cílením do oblasti genů nestojí mechanismy přímo spojené s transkripcí, ale nějaký jiný motiv vyskytující se v místech genů (ENGELMAN 2005, MAXFIELD et al. 2005), což se u RSV zatím zdá možné, nebo užitím jiných laboratorních technik a rozdílných buněčných kultur (BUSHMAN et al. 2005).

U hojně zkoumaného HIV-1 i ostatních lentivirů je situace o to složitější, že je zde integrace cílena do transkripčně aktivních oblastí pomocí buněčného proteinu.

Pokud ale dojde k inhibici tohoto mechanismu, sníží se počet integrací na 10% a zůstane pouze lehká preference pro oblasti genů, podobná RSV (SHUN et al. 2007). U MLV nebyl vliv transkripční aktivity na integraci zatím prozkoumán, ale u rodově příbuzného člena gammaretrovirů PERV (Porcine endogenous retrovirus) byl na lidských HEK-293 buňkách pozorován zvyšující se počet integrací spolu se stoupající hladinou exprese, s vyjímkou genů s nejvyšší úrovní exprese (MOALIC et al. 2006).

Celkově se pak dá říci, že nehledě na konkrétní preference různých retrovirů jsou transkribované oblasti retroviry preferovány (pravděpodobně za to vděčí spíše snazší přístupnosti, než transkripční aktivitě samé), ale že příliš vysoká hladina exprese může integraci naopak inhibovat. U rodu gammaretrovirus je pak transkripční aktivita důležitá kvůli jeho cílení do oblastí promotorů, kde se předpokládá interakce s transkripčními faktory (ENGELMAN 2005).

S transkripční aktivitou cílové DNA také úzce souvisí její metylační status. V této souvislosti se jeví jako zajímavá schopnost RSV překonat přechodně metylační blok v hostitelské DNA, kdy HEJNAR et al. (2003) pozorovali po integraci do původně hypermetylované části DNA demetylaci tohoto úseku a úspěšnou transformaci těchto buněk onkogenem *v-src*. Tento stav ale trval jen přechodně a metylace byla po čase obnovena, včetně metylace provirové DNA.

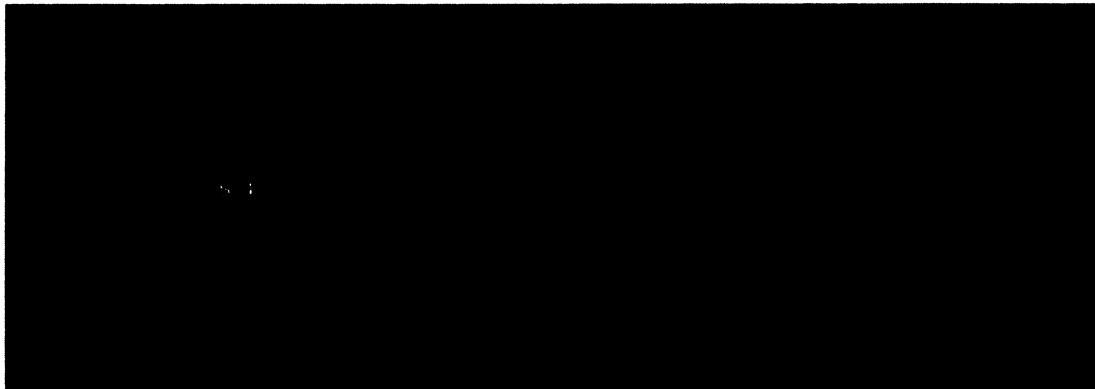
### **3.5.4. Buněčný cyklus**

Zde může hrát roli zejména pozice a stav chromozomů během různých fází buněčného cyklu a může být i jednou z příčin charakteristického integračního vzorce u MLV, který potřebuje pro svou integraci, aby cílová buňka prošla mitózou (ROE et al. 1993). Také může mít vliv usnadnění přístupu k DNA během replikace. A pozice chromozomu v jádře pak hypoteticky i na preferenci konkrétního chromozomu (BUSHMAN et al. 2005). Zatím ale není žádný jasný důkaz, že by buněčný cyklus hrál důležitou roli v integračním cílení retrovirů.

### **3.5.5. Buněčné proteiny schopné vázat PIC**

Vliv buněčných proteinů na místo integrace byl poprvé popsán u kvasinkových LTR retrotransposonů. LTR retrotransponzy jsou retrovirům příbuzné retroelementy, které si kódují vlastní integrázu, reverzní transkriptázu a mají také LTR, ale postrádají gen pro obalový protein a jsou proto odkázány na jednu buňku (BOEKE a DIVINE 1998). To činí výběr integračního místa zvlášť důležitým, jelikož integrace do špatného místa může znamenat smrt hostitelské buňky a tím i retrotransponzu. Retroelementy kvasinek mají proto vyvinutou speciální strategii, spoléhající na navázaní pro jednotlivé druhy specifických proteinů, které jim interakcí s cílovou DNA umožní integraci do konkrétní oblasti v hostitelské DNA (ZHU et al. 2003).

Nejdetajněji je tento proces popsán u retrotransponz Ty3 a Ty5. Pro Ty3 je charakteristické místo integrace ležící 1-4 bp před transkripčním počátkem tRNA genů, do kterého se integruje díky interakci preintegračního komplexu s transkripčními faktory TFIIIB/TFIIIC polymerázy III (KIRCHNER et al. 1995). Ty5 je naopak směrován do transkripčně neaktivních oblastí genomu, jako jsou telomery. To je umožněno interakcí virové integrázy s transkripcí umělujícím proteinem Sir4, který se váže na DNA právě v transkripčně neaktivních oblastech genomu (XIE et al. 2001) (obr. 7). Za vznikem tohoto mechanismu pak s velkou pravděpodobností stojí molekulární mimimkry, kdy BRADY et al. (2007) odhalili, že buněčný protein Esc1 interahuje se stejnou doménou na Sir4 jako Ty5 integráza a že 75% mutací rušících IN-Sir4 interakci úspěšně zabránil i interakci IN-Esc1.



Obrázek 7. Schéma znázorňující DNA vázajícími proteiny zprostředkovanou integraci ve specifické oblasti u LTR retrotransposonů.  
Převzato od Wu a Burghes (2004)

Za vznikem takovýchto integračních mechanismů a preferencí pro oblasti heterochromatinu, telomer, či promotorové oblasti genů přepisovaných DNA polymerázou III, pak nejspíše stál selekční tlak a dokresluje důležitost místa integrace v kontextu s životní strategií konkrétního retrovиру, či retrotransponoru.

Proteiny s podobnou funkcí byly zanedlouho identifikovány i u retrovirů, a nejdetajněji byly zatím popsány u HIV. Integrační vzorec HIV-1 i ostatních lentivirů se vyznačuje extrémně silnou preferencí pro aktivní geny (ELLEDER et al. (2002), SCHROEDER et al. (2002)), a bylo již popsáno několik virových i buněčných proteinů interagujících s jeho preintegračním komplexem. (VANDEGRAAFF a ENGELMAN 2007)

Jako nejdůležitější se jeví buněčný protein LEDGF/p75, kdy po snížení jeho exprese, byla výrazně snížena preference HIV-1 pro transkripční jednotky v několika typech buněčných linií. Jistá preference však přetrvala a není zřejmé, zda-li to zapříčinila zbytková exprese, nebo jestli je ve hře ještě nějaký další faktor (CIUFFI et al. 2005). Vliv zbytkové exprese pak vyloučila knockout studie kterou vykonal SHUN et al. (2007) a která zároveň i potvrdila, že LEDGF/75 není pro integraci samotnou nezbytný. Jedinečnost tohoto mechanismu a jeho závislost na lentivirální integráze prokázal LEWINSKI et al. (2006), když vytvořil chimerický HIV-1-MLV virus, který nesl gen pro MLV integrázu a jehož integrační vzorec byl shodný s MLV.

Mezi další zúčastněné, ale doposavad méně prozkoumané a nejspíše i z pohledu integračního cílení méně nedůležité, patří BAF (LEE a CRAIGIE 1998) a enzym p300 (CERESETO et al. 2005). U BAF byly zjištěny interakce s PIC MLV i

HIV a u HIV byl nalezen i uvnitř virionů. Jeho hlavní přínos pro retrovirus tkví v ochraně PIC před autointegrací (SEGURA-TOTEN, WILSON 2004).

Histon acetyl-transferáza p300 pak acetylací lizinů na C-terminální doméně integrázy zvyšuje účinnost integrace, ale nejspíše také nemá vliv na její cílení.

Velmi zajímavý pokus, dokazující důležitost interakcí uskutečnil TAN et al. (2004), když sestrojili umělý fúzní protein sestavený z HIV-1 integrázy a syntetického polydactyl zinc finger proteinu E2C. Pročistěný fúzní protein pak *in vitro* dokázal ovlivnit integraci virové DNA do blízkosti 18bp-E2C charakteristické sekvence. Pokus pak ještě vylepšili provedením *in vivo*, kdy sestrojili infekční viriony s modifikovanou integrázou, která nesla E2C protein na N nebo C konci. Po infekci porovnali integrační místa virů s divokým a umělým typem integrázy zjistili, že integráza s E2C na C konci integrovala virus sedmkrát častěji v blízkosti E2C vázající oblasti a integráza s E2C na N konci pak dokonce desetkrát častěji (TAN et al. 2006).

Tyto výsledky ukazují, že DNA vazebné proteiny jsou u některých retrovirů a retrotransponů jedním z hlavních hráčů v integračním cílení a jsou příslibem do budoucna, kdy by mohly posloužit ke konkrétnímu cílení retrovirových vektorů a umožnit tak bezpečnější genovou terapii.

### **3.6. Integrační preference u různých skupin retrovirů**

Během několika posledních let byly rozsáhlými *in vivo* výzkumy charakterizovány integrační preference u prakticky všech významných retrovirů, adenovirů, retrotransponů, a ostatních mobilních elementů. A velmi rozdílné integrační schéma napříč zkoumaným spektrem dokládá, že integrační preference jsou rodově specifické a odráží se v nich životní strategie jednotlivých skupin.

První retrovirus u kterého byla *in-vivo* v dostatečném počtu provedena charakteristika integračních míst byl HIV-1 (rod *Lentivirus*). Práci uskutečnil ELLEDER et al. (2002) tak, že mapovali 48 dříve publikovaných integračních míst viru HIV, s právě uveřejněným lidským genomem. Výsledek byl velmi překvapivý, ukázalo se totiž, že virus zcela průkazně preferuje oblasti s vysokou hustotou genů (54% oproti 22% značícím náhodnost integrace) a s nimi související GC-bohaté

oblasti a cytogenetické proužky G. Tyto výsledky později potvrdil výzkum provedený s 524 integračními místy v genomu buněčné linie SupT1. Navíc bylo zjištěno, že HIV silně preferuje aktivní geny (69% oproti 33%) po celé délce a sám k jejich aktivaci přispívá a nalezli také hotspot na lokusu chromozomu 11q13 (obr.8) (SCHROEDER et al. 2002).



**Obrázek 8.** Přehled integrací HIV-1 cDNA do lidského genomu.  
Místa integrací jsou zobrazena jako "balónky" nad lineárními chromozomy. Fialové indikují HIV-1; červené od HIV odvozený vektor; a zelené *in vitro* kontrolu. Barevná čára ve vrchní části vyobrazených chromozomů značí integraci do genu (zlatá), nebo mimo gen (šedá). Spodní část pak ukazuje relativní genovou hustotu na chromozomu - regiony s větší koncentrací genů jsou znázorněny červeněji. Centromery jsou vyznačeny šedými obdélníky v horní části. Chromozom Y není v použité buněčné linii SupT1 přítomen. (Schroeder et al. 2002)

Další práce uskutečnil na HeLa a H9 lidských buňkách WU et al. (2003) a zahrnovala krom HIV-1 i *gammaretrovirus* MLV. Nebyl zde potvrzen hotspot pozorovaný Schroeder et al., možná kvůli volbě jiných buněk, nebo kvůli nižšímu počtu sekvenovaných integračních míst (379), ale jinak údaje souhlasily. MLV pak zcela průkazně preferoval pro svou integraci místa transkripčních počátků a přilehlé oblasti, když v oblasti 5 kb okolo transkripčních počátků proběhlo 20% integrací z 903 a náhodná integrace měla nastat pouze ve 4% případů. Tento výsledek pak

naznačuje, že v případě leukémií dětských pacientů při X-SCID genové terapii (HACEIN-BEY-ABINA et al. 2002) mohla hrát úlohu integrační preference MLV. Totožných výsledků dosáhl poté i HEMATTI et al. (2004), když porovnali integrační preference SIV a MLV v opičích HSC (hematopoietic stem cell). Tyto výsledky ukázaly, že rodově příbuzné retroviry, s podobnou integrázou, jako jsou výše uvedené lentiviry HIV-1 a SIV sdílejí podobný, nebo dokonce shodný integrační mechanismus. Další práce provedené na rodově příbuzných retrovirech, kdy SIV i v lidských buňkách opět vykazoval shodný integrační vzorec s HIV-1 (CRISE et al. 2005) a gammaretrovirus PERV s MLV (MOALIC et al. 2006), pak tuto teorii dále podpořily.

HIV-2 *in vitro* vykazoval stejný integrační vzorec jako HIV-1, ale u integrací získaných z infikovaných osob vykazoval větší preferenci pro heterochromatin, což může být jednou z příčin jeho menší patogenicity (MACNEIL et al. 2006).

Výsledky integračních preferencí FV (Foamy virus), zástupce rodu *spumavirus* v lidských buňkách zveřejnili téměř v totožném čase 2 skupiny a i přes rozdílnou metodiku a použití jiných buněčných kultur dosáhli totožných výsledků. FV na rozdíl od ostatních retrovirů nevykazuje preferenci pro geny, ale podobně jako MLV upřednostňuje CpG ostrovy a oblasti promotorů, ovšem jen s poloviční četností (NOWROUZI et al. 2006, TROBRIDGE et al. 2006).

MMLV (mouse mammary tumor virus) byl prvním z rodu *betaretrovirus*, u kterého byly prozkoumány integrační preference. Práce byla provedena na myších i lidských buněčných liniích a neodhalila žádnou preferenci. MMLV tak zatím jako jediný vykazuje náhodný integrační vzorec. Je ale možné, i když nepravděpodobné, že použité buňky neprodukují buněčné faktory cílící tento virus (FASCHINGER et al. 2008).

Z rodu *deltaretrovirus* byl nejlépe charakterizován integrační vzorec HTLV-1 (Human T-cell leukemia virus type 1). První práce byly většinou provedeny na vzorcích z chronických ATLL (adult T-cell leukemia/lymphoma) pacientů a nacházely preferenci pro aktivně transkribované oblasti (OZAWA et al. 2004), poté ale DOI et al. (2005) ukázal, že za touto preferencí stojí selekční tlak v leukemických buňkách po integraci. Neovlivněnou preferenci tudíž odhalila až práce DERSE et al. (2007), kdy použili vektor, který nekódoval virové proteiny. Odhalili integrační

vzorec velmi podobný ASLV, kdy byly velmi lehce preferovány geny, transkripční počátky a CpG ostrovky.

ASLV současně zkoumaly dvě skupiny. NAREZKINA et al. (2004) osekvenovali 226 integračních míst v HeLa buňkách a zjistili jen slabou, přibližně dvounásobnou preferenci pro aktivní geny, zvýšení exprese nezaznamenali. K podobnému výsledku dospěla i další skupina, která charakterizovala 469 integračních míst ASLV podskupiny A v 239T buňkách. Navíc ještě charakterizovali další integrační místa MLV a HIV na jiných buněčných liniích, a podpořili jimi výsledky předchozích studií (MITCHELL et al. 2004). Na kuřecích fibroblastech pak potvrdil výsledky předchozích skupin s HIV-1 a s podskupinou A-ASLV BARR et al. (2005) a s RSV podskupiny C REINIŠOVÁ et al. (2008). Navíc pak obě skupiny pozorovali u ptačích virů sníženou integraci do mikrochromozomů, což bylo proti očekávání, jelikož mikrochromozomy jsou bohaté na GC – preferovaný cíl těchto virů. U HIV-1 tomu bylo naopak, mikrochromozomy byly dle očekávání upřednostňovány (BARR et al. 2005).

### **3.7. Retrovirové vektory**

Od retrovirů odvozené vektory se využívají především ve 3 druzích aplikacích, které ovšem kladou na použitý vektor zcela odlišné požadavky. Hlavní nevýhodou, která je společná všem retrovirům kromě lentivirů je nutnost mitotického dělení buňky pro účinnou integraci.

V případě genové terapie je hlavním požadavkem bezpečnost vektoru, zejména co se rizika inzerční mutageneze týče. Proto je důležité, aby použitý retrovirus co nejméně preferoval rizikové oblasti, jako jsou geny a jejich promotory. (MACHENO-CORVO, MARTIN-DUQUE 2006).

V případech, kdy je retrovirus aplikován jako genová past (gene trap) je naopak prioritou změna transkripční aktivity zasaženého genu a tím umožnění studia jeho funkce. Toho lze nejlépe dosáhnout vektory s preferencí pro oblasti genů a se silnými transkripčními aktivátory (HACKETT et al. 2007).

Shodné požadavky jsou kladený i na vektory které mají sloužit pro expresi rekombinantních proteinů.

## **4. MATERIÁL A METODY**

### **4.1. Práce s buněčnými kulturami a viry**

#### **4.1.1. Použité buněčné linie**

**Pakážovací linie E6** - odvozená od permanentní buněčné linie kuřecích fibroblastů DF-1. Ta umožňuje replikaci ASLV podskupin A, B, C a neobsahuje endogenní viry příbuzné ASLV (SCHAEFER-KLEIN et al. 1998) . V E6 je navíc stabilně integrován expresní vektor pcGagPol (viz. kapitola 4.3.12.) exprimující geny ASLV *gag* a *pol*. Na počátku genu *gag* se nachází transkripční enhancer, jenž zajišťuje účinnou expresi strukturních genů a pomáhá tím dosáhnout dostatečného titru viru.

#### **4.1.2. Viry**

**Pseudotypovaný virus Rousova sarkomu, podskupiny C** – Replikačně kompetentní retrovírový vektor, jehož základem je LTR--v-src--LTR konstrukt pH19FM3, zabalený do obalového glykoproteinu G viru vezikulární stomatitidy (VSV-G) a využívající strukturní proteiny, polymerázu a integrázu kódovanou pcGagPol plazmidem z pakážovací line E6. Obalový G glykoprotein nepotřebuje pro vstup do buňky žádný specifický receptor. Použitá kuřecí linie P je vůči ASLV podskupině C semirezistentní, pseudotypizací viru byl problém rezistence odstraněn.

#### **4.1.3. Používaná média**

Pro kultivaci kuřecích fibroblastů jsme použili medium Eagle's H-MEM (připraveno v UMG AVČR) s 1% NaHCO<sub>3</sub> doplněné:

MEM směsí neesenciálních aminokyselin (Sigma)  
pyruvátem sodným (0,12 g/l)  
10% telecího séra  
roztoky antibiotik

Pro transfekci Lipofectinem 2000 (Invotrogen) jsme použili medium opti-MEM bez antibiotik.

#### **4.1.4. Kultivace buněk, rozmražování buněk**

Připravili jsme na Petriho misku médium pro příslušné buňky. Buňky jsme vyseli po  $1,5 \times 10^6$  na petriho misku o průměru 10 cm (P100). Ampuli vytaženou z dusíku jsme ponořili do vodní lázně o 37°C. Po rozmražení jsme buňky přenesli do zkumavky s 1 ml média a centrifugovali 10 minut při 174 x g. Buněčný pelet jsme resuspendovali v médiu a přidali do Petriho misky.

Buňky byly pasážovány 1-2 x týdně. Po odsátí média a opláchnutí PBS jsme buňky rozvolnili směsí 0,5% trypsinu s 0,02% EDTA v poměru 1:1 (0,5 ml na P60, 1 ml na P100) a nechali 10 minut ve 37°C. Část buněk jsme po resuspendaci v médiu vyseli na misky.

#### **4.1.5. Sklizení buněk**

Po odsátí média jsme buňky opláchlí PBS a rozvolnili směsí 0,5% trypsinu s 0,02% EDTA v poměru 1:1 (3 ml na P60, 6 ml na P100) a nechali ve 37°C 10-15 min. Resuspendované buňky jsme centrifugovali (15 min, 1000 x g, 4°C).

#### **4.1.6. Transfekce buněk**

Buňky jsme transfekovali pomocí Lipofectaminu 2000 (Invitrogen). Den předem jsme vyseli buňky E6 na petriho misky P100 v množství  $1,5 \times 10^6$ -  $3 \times 10^6$ , aby v době transfekce byly v 90-95% hustotě. Další den jsme vyměnili médium a po hodině jsme zahájili transfekci.

Ta se prováděla jako kontransfekce vektory pH19FM3 a VSV-G. Nejprve jsme smíchali DNA obou vektorů v poměru 6:1 tak, aby celkové množství DNA bylo 28  $\mu$ g a naředili tento mix v 1,5 ml opti-MEM media. Poté jsme promíchali Lipofectamine 2000 a naředili jej 1,5 ml opti-MEM media a nechali inkubovat při pokojové teplotě (RT, room temperature) po dobu 5 minut. Poté jsme oba naředěné roztoky slili dohromady, lehce promíchali a nechali 20 minut inkubovat při RT. Mix byl poté přidán k 24 h kultuře a po 6 hodinách bylo vyměněno médium za selektivní. Virus byl sebrán po 24 hodinách.

#### **4.1.7. Sběr a zamražení viru**

Večer před odběrem jsme transfekovaným buňkám vyměnili médium. Druhý den ráno jsme pak 24 hodin po infekci odebrali médium a doplnili jej telecím sérem na finální obsah 10% séra, to jsme poté centrifugovali 10 minut při 150 g, 4°C a supernatant obsahující virové partikule sme přefiltrovali přes 0,45  $\mu$ m filt (Whatman).

Přefiltrované médium s virovými partikulemi sme poté rozdělili do 1 ml konzervačních zkumavek a zamrazili v -80°C.

#### **4.1.8. Titrace viru**

Titr viru byl určen fokálním testem (SVOBODA et al. 1968). Nejdříve jsme připravili řadu médií s virem od  $10^0$  do  $10^{-6}$ . Poté jsme od každé koncentrace použili 100  $\mu$ l ředěného viru na infekci CEF vysetých na petriho misku P60. Od každého ředění byly takto infikovány dvě misky P60. Fokusy byly odečítány v období 10 až 15 dnů po infekci. Titr viru byl vyjádřen počtem FFU/ml (focus forming units).

#### **4.1.9. Inokulace kuřat**

Pro inokulaci VSV-G pseudotypizovaným virem RSV byla zvolena kuřata linie P, jenž postrádají endogenní retroviry produkující falešné pozadí při analýze integračních míst. Inokulace byla provedena formou injekce 0,5 ml viru do prsního svalu, nebo do podkoží křídelní řasy. Titr použitého viru byl 100 - 1000 FFU/ml.

Růst nádorů do požadované velikosti trval 4-5 týdnů, poté byla kuřata utracena a nádorová tkáň odebrána pro izolaci chromozomální DNA.

### **4.2. Materiál a metody, práce s nukleovými kyselinami**

#### **4.2.1 Použité vektory**

**pGEM T-easy vektor** dsDNA vektor s 1-nukleotidovými T-přesahy pro AT klonování produktu PCR (viz kap.x) v polyklonovací oblasti. Vektor obsahuje operon lacZ, který umožnuje barevnou selekci pro klony pozitivní na inzerovaný fragment, gen pro ampicilinovou rezistenci, promotory SP6 a T7 a vazebná místa pro M13 primery *Forward* a *Reverse*. (Promega)

**pH19FM3** dsDNA vektor nesoucí LTR—v-src—LTR a gen pro ampicilinovou rezistenci, přičemž v 5' LTR se nachází vnesená restrikční místa *MmeI* a *FseI*.

Odvozen od vektoru pH19KE

**pVSV-G** pakážovací vektor od firmy Clontech. Kotransfekován s plazmidem pH19FM3. Použit k pseudotypizaci viru pomocí VSV-G glykoproteinu (viz. kapitola 4.1.2.)

#### **4.2.2. Izolace chromozomální DNA**

Nejprve jsme buněčnou kulturu ( $2\text{-}10 \times 10^6$  buněk) popřípadě tkáň (1 g) lzyovali 5-12 h v 800  $\Omega$  lyzačního pufru (SDS 1%, EDTA 250 mM, proteináza K 1 mg/ml) za stálého otáčení při 55°C. Následně jsme přidali do každé zkumavky 10  $\Omega$  Rnázy A (0,3 mg/ml) a dali inkubovat na 1 h do 37°C za stálého pomalého rotačního promíchávání

K izolaci chromozomální DNA jsme použili 200  $\Omega$  roztoku fenolu a chloroformu (1:1), který rozdělí buněčný lyzát na lipidy, které zůstanou ve fenolové fázi, proteiny, které se vysrážejí na mezifázi a nukleové kyseliny a ionty, které zůstanou ve vodní fázi.

Centrifugovali jsme 5 minut, při 21000 x g a RT. Odebrali jsme vazkou vodní fázi a proces opakovali, dokud jsme dokonale neoddělili DNA od proteinů.

DNA jsme vysráželi ve shodném množství 96% ethanolu, lehce zvortexovali a stáčeli 4 minut při 4°C a maximálních otáčkách na centrifuze. Pelet DNA jsme opláchli 80% vychlazeným etanolem, 5 minut stáčeli při maximálních otáčkách a nechali oschnout. DNA jsme rozpustili v 50  $\Omega$  T<sub>10</sub>E<sub>0,1</sub> (Tris, pH8 10 mM; EDTA 0,1 mM).

Koncentraci DNA jsme zjišťovali pomocí spektrofotometru a ověřovali pomocí elektroforézy.

#### **4.2.3. I-PCR**

Pro detekci a amplifikaci sekvencí na rozhraní proviru a chromozomální DNA jsem použili metodu inverzního-PCR (ANGLEY et al. 1994).

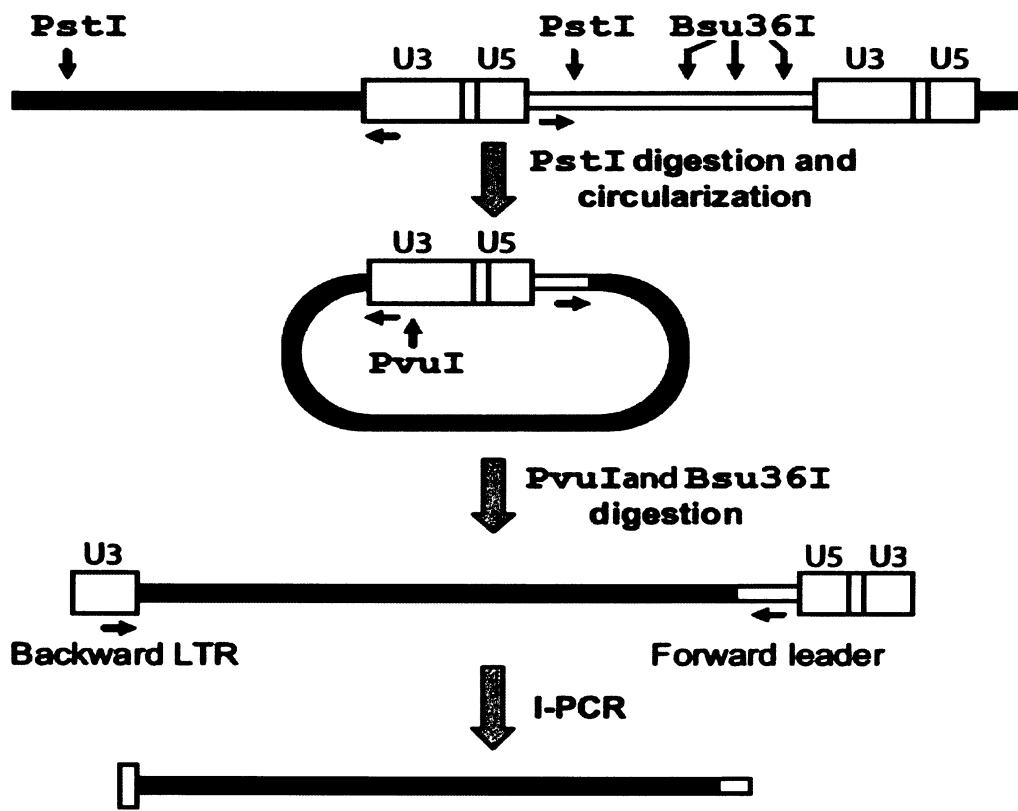
##### **4.2.3.1. Princip I-PCR**

I-PCR je metoda, která používá od sebe cílené primery, jenž se po digesci a selfligaci ocitnou v protisměrné poloze nezbytné pro úspěšnou reakci (obr. 9).

Prvním krokem je rozštěpení chromozomální DNA obsahující provirus restrikčním enzymem, který štěpí jednou za cca 500 bp v chromozomální DNA a

jednou v oblasti leader sekvenci proviru, za námi navrženým primerem. My jsme použili restriktázu *PstI*. V restrikční směsi se nyní vyskytují i požadované fragmenty složené z provirové i kuřecí DNA, které v dalším kroku cirkularizujeme selfligací. Tím vznikne kruhová molekula DNA, která obsahuje provirový 5'LTR s kusem leader sekvence a část upstream kuřecí DNA o neznámé délce. V posledním kroku použijeme restriktázu štěpící v LTR mezi našimi primery, čímž vznikne fragment s korektní orientací primerů. Tuto digesci provedeme spolu s enzymem mnohonásobně štěpícím v kódující sekvenci proviru. Ten poslouží k destrukci vnitřních provirových fragmentů s 3' LTR, které by mohly interferovat v PCR reakci a zvyšuje tak její výtěžek. My použili enzymy *PvuI* a *Bsu36I*.

Nakonec provedeme klasickou PCR reakci. Výsledným produktem je fragment kuřecí DNA ohraničený leader sekvencí z jedné a začátkem U3 sekvence 5' LTR z druhé strany.



9. Schéma i-PCR: Plná čára je kuřecí DNA, prázdná značí provirus. Horizontální šipky jsou primery, jeden je v U3 oblasti, druhý v místě forward leader. Vertikální šipky jsou místa štěpení.  
Převzato a upraveno od Reinišová et al. (2008)

#### **4.2.3.2. Postup**

4-6  $\mu$ g chromozomální DNA jsme štěpili restrikčním enzymem *Pst*I (NEB, 5U) v 50  $\mu$ l reakci při 37°C přes noc. Přibližně 300 ng jsme nanesli jako kontrolu na elektroforetický gel.

Dalším a ještě několikrát použitým krokem byla precipitace DNA fenol-chloroformem. Připravili jsme si srážecí roztok: 35  $\mu$ l 5M Na AC, 3  $\mu$ g glykogenu a doplnit vodou do objemu 300  $\mu$ l, a ten přidali k DNA. Poté jsme v digestori přidali 250  $\mu$ l roztoku 1:1 fenol-chloroform pH 8, protřepali a stočili 2 minuty při 18000 x g a RT. Supernatant jsme odebrali do nových zkumavek a srazili DNA 200  $\mu$ l izopropanolu. Protřepali jsme a nechali stát 5 minut v RT. Centrifugovali jsme 15 minut, při 22000 x g a 15°C. Následoval oplach 80% RT ethanolem a 5 minut centrifugace při 15°C a maximálních otáčkách. Vzniklý pelet jsme vysušili a rozpustili ve 20  $\mu$ l H<sub>2</sub>O.

Koncentrace DNA byla změřena na spektrofotometru a 2-3  $\mu$ g byly přidány do selfligační reakce. Ta se kvůli minimalizaci vzájemných ligací různých fragmentů přes *Pst*I místo dělala v objemu 100  $\mu$ l. Bylo použito 400 jednotek T4 DNA ligázy, 10  $\mu$ l 10 x T4 DNA pufru a doplněno H<sub>2</sub>O do finálního objemu 100  $\mu$ l. Reakce proběhla při 15°C přes noc.

Opět následovala precipitace DNA fenol-chloroformem. Po přečištění a rozpuštění v H<sub>2</sub>O jsme DNA rozštěpili společně enzymy *Pvu*I (NEB) a *Bsu*36I (NEB) v celkovém objemu 50  $\mu$ l. Reakce proběhla při 37°C přes noc. Reakční směs sme poté opět precipitovali fenol-chloroformem, DNA rozpustili ve 20  $\mu$ l vody a změřili její koncentraci na spektrofotometru.

150-200 ng DNA bylo použito do i-PCR reakce, která proběhla v objemu 20  $\mu$ l.

Složení reakční směsi:

DNA .....	150-200 ng
<i>Taq</i> pol. (5U/ $\mu$ l, Roche).....	0,1 $\mu$ l
10X pufr (s 20 mM MgCl <sub>2</sub> ).....	2 $\mu$ l
DMSO 2,6% +Betain 2,6 M	7 $\mu$ l
směs dNTP (10 mM).....	0,4 $\mu$ l
Primery FW a R (10 mM).....	0,2 + 0,2 $\mu$ l
H <sub>2</sub> O.....	doplnit do 20 $\mu$ l

Pro amplifikaci byly použity tyto primery (tab. 4.1):

Název primeru	sekvence primeru
2 - Forward leader	5'CCTCATCCGTCTCGCTTATCG3'
1 - LTR	5'CCTTACTACCACCAATCGGCA3'

Amplifikace probíhala v termocykleru MJ Research DNA Engine DYAD za následujících podmínek (tab. 4.2.) :

	teplota	doba trvání	počet cyklů
úvodní denaturace	95°C	3 minuty	
denaturace	94°C	20 sekund	
nasednutí primerů	58°C	50 sekund	34
DNA polymerace	72°C	2 minuty	
závěrečná polymerace	72°C	3 minuty	

I-PCR produkty jsme nanesli na 1% agarózový gel. Naším cílem byly fragmenty o velikosti minimálně 250 bp. V případě jednoho pozitivního fragmentu jej stačilo pouze přečistit produktem EXOSapIt a osekvenovat. Pokud bylo fragmentů z jedné reakce více, použili jsme buď preparační gel a poté jednotlivé fragmenty vyřezali, přečistili a osekvenovali, nebo ligaci do vektoru pGEM T-easy a poté transformaci do bakterií,  $\beta$ -gal selekci, koloniové PCR, přečištění a sekvenaci.

#### 4.2.4. Izolace PCR produktu z gelu

Abychom se zbavili různých vedlejších nebo poškozených produktů PCR reakce, rozdělili jsme DNA na 1% ultračistém agarózovém gelu (ultraPURE agarose, GibcoBRL) s ethidium bromidem. PCR produkt požadované velikosti jsme vyřízli skalpelem pod UV lampou ( $\lambda = 366$  nm). Z gelu jsme DNA izolovali pomocí QIAEX II Gel Extraction Kit (150, QIAGEN) podle návodu přiloženého výrobcem. DNA jsme elulovali ve 20  $\mu$ l H<sub>2</sub>O.

#### **4.2.5 Ligace PCR produktu s místem integrace do vektoru pGEM-T Easy**

*Taq* polymeráza přidává samovolně na 3' konce PCR produktů adenin. Toho je využito u vektoru pGEM-T Easy (Promega) s přesahy tyminu na 3' konci. Tuto metodu jsme použili pro vložení získaných fragmentů do vektoru.

Ligační směs:

pGEM-Teasy vektor 50ng/ l.....	1 1
PCR produkt.....	3 1
2 x ligační pufr (Promega).....	5 1
T4 ligáza (U3) (Promega).....	1 1

Směs jsme nechali přes noc při 4°C. Následně jsme ji použili pro transformaci kompetentních bakterií *Esherichia coli* XL-1 Blue MRF'.

#### **4.2.6. Příprava kompetentních bakterií**

Pro přípravu kompetentních bakterií byla použita rubidium chloridová chemická metoda a bakteriální kmeny *E. coli* XL-1 Blue MRF'. Do 5 ml média LB byla zaočkována rozmražená bakteriální kultura a nechala se narůst přes noc při 37°C za stálého třepání. Následující den byl 1 ml z této kultury zaočkován do 100 ml LB média a bakterie se nechaly dorůst za intenzivního třepání při 37°C do hustoty odpovídající optické densitě 0,4 - 0,5 (měřeno při  $\lambda = 590$  nm). Po 10-15 minutách na ledu byla bakteriální kultura centrifugována (7 min, 850 g, 4°C). Následně byl bakteriální pelet resuspendován ve 30 ml vychlazeného roztoku TFB I (100 mM RbCl, 50 mM MnCl<sub>2</sub>, 10 mM CaCl<sub>2</sub>, 30 mM octan draselný, 15% glycerol, pH 5,8 upraveno pomocí kyseliny octové; roztok sterilizován filtrací přes filtr 0,22 μm a skladován ve 4°C). Po dalších 20 minutách inkubace na ledu byla suspenze bakterií opět centrifugována (7 min, 850 g, 4°C). Pelet byl resuspendován ve 4 ml zmražovacího roztoku TFB II (10mM MOPS pH 7, 10 mM RbCl, 75 mM CaCl<sub>2</sub>, 15% glycerol, pH 6,8 upraveno pomocí NaOH; roztok sterilizován filtrací a skladován v -20°C). Takto vzniklá suspenze kompetentních bakterií byla na ledu rozdělena po 100 μl do zkumavek. Zkumavky pak byly přemístěny do směsi suchého ledu a ethanolu a po 20 min uskladněny v -80°C.

#### **4.2.7. Transformace bakterií**

Do zkumavky s kompetentními bakteriemi jsme ještě před rozmrznutím přidali připravenou ligační směs a inkubovali 30 min na ledu. Pak jsme bakterie dali na 1,5 minuty do vodní lázně o teplotě 42°C. Teplotním šokem jsme zvýšili účinnost transfekce. Bakterie jsme přenesli do 0,9 ml LB média a nechali 45 min třepat při 37°C. Posléze jsme část suspenze rozetřeli na bakteriální LB plotnu (1,5% agar; po vychladnutí na 50°C je přidán ampicilin 100 µg/ml, chromogenní substrát S-gal (0,3 mg/ml, Sigma), induktor  $\beta$ -galaktosidázy izopropyl- $\beta$ -D-1-thiogalaktopyranosid 0,03 mg/ml, citran železito-amonný 0,5 mg/ml; na petriho misky P100 po 10-15 ml) Po 14-20 hodinách narostly kolonie bakterií, z nichž jsme vybrali 2-5 bílých, zaočkovaly do 2 ml LB média s ampicilinem (100 µg/ml) a nechali narůst přes noc při 37°C za stálého třepání.

#### **4.2.8. Koloniové PCR**

Do 0,2 ml zkumavky jsme přidali 5 µl H<sub>2</sub>O a špičkou do ní přenesli část analyzované kolonie. Část kolonie jsme přenesli na bakteriální LB plotnu s 1,5% agarem obsahující ampicilin (100 µg/ml) s nakreslenou mřížkou s očíslovanými okénky, přičemž každé kolonii odpovídá jedno okénko. Zkumavky jsem dal na 3 min do 98°C. Na ledu jsme přidali:

Pufr 10 x.....	2 µl
Primer 1 Reverse.....	0,2 µM
Primer 2 Forward.....	0,2 µM
dNTP.....	200 µM
<i>taq</i> polymeráza.....	0,3 U
H <sub>2</sub> O.....	doplnit do 16 µl

Pro amplifikaci byly použity tyto primery (tab. 4.3.):

Název primeru	sekvence primeru
1 - Forward leader	CCTCATCCGTCTCGCTTATTCG
2 - LTR	CCTTACTACCACCAATGGCA

Amplifikace probíhala v termocykleru MJ Research DNA Engine DYAD za následujících podmínek (tab.4.4.):

	Teplota	Doba trvání	Počet cyklů
Denaturace	96°C	15 s	25
Nasednutí primerů	60°C	45 s	
DNA polymerace	60°C	45 s	

PCR reakci jsme analyzovali na agarózovém gelu. Z reakcí, ze kterých jsme získali 1 kvalitní PCR produkt odpovídající délky, jsme odebrali 30 µg DNA a tu přečistili pomocí EXOSapIT roztoku a sekvenovali. Kolonii s požadovaným fragmentem jsme také mohli zaočkovat do 2 ml LB média, nechat přes noc třepat při 37°C a následně provést minipreparaci plazmidové DNA.

#### 4.2.9. Sekvenace DNA

Čistou DNA jsme sekvenovali pomocí komerčního kitu AmpliTaq®FS L' Big Dye Terminator obsahující směs terminálních dideoxinukleotidů značených různou fluorescenční značkou, pufr a DNA polymerázu (PREMIX).

Sekvenace fragmentů s integračním místem:

Při velikosti fragmentu do 600 bp jsme použili pouze primer 2 - forward, v případě většího fragmentu jsme udělali i druhou reakci s primerem 1 - reverse (viz tab.4.6.).

Pro sekvenaci plazmidu pH19FM3 jsme použili univerzálních primerů (tab 4.5.):

Název primeru	Sekvence primeru
pUC/M13 Reverse	TCA CAC AGG AAA CAG CT
pUC/M13 Forward	GTA AAA CGA CGG CCA GT

Pro sekvenaci integračních míst jsme použili primery pro I-PCR (tab 4.6.):

Název primeru	sekvence primeru
2 - Forward leader	CCTCATCCGTCTCGCTTATTAG
1 - Reverse LTR	CCTTACTACCACCAATCGGCA

Složení sekvenační směsi:

DNA templát.....	1-2 0g
primer (10uM).....	1 0l
PREMIX.....	2 0l
H <sub>2</sub> O.....	doplnit do 10 0l

Amplifikace probíhala v termocykleru MJ Research DNA Engine DYAD za následujících podmínek (tab. 4.7.):

	Teplota	Doba trvání	Počet cyklů
Úvodní denaturace	96°C	30 sekund	
Denaturace	96°C	10 sekund	25
Nasednutí primerů	58°C	5 sekund	
DNA polymerace	60°C	4 minuty	

Výsledný produkt jsme naředili 80 0l H<sub>2</sub>O a 10 0l 3 M NaAc (pH 5,2) a precipitovali 80 0l izopropanolu. Vzorky jsme centrifugovali (30 min, 18000 g), pelet DNA opláchli 250 0l 70% ethanolu a opět centrifugovali 5 minut. Vysušený pelet jsme rozpustili v 25 0l Template Supression Reagent (TSR, Applied Biosystems). Po 2 min denaturaci v 90°C proběhla samotná sekvenace pomocí kapilárního sekvenátoru ABI PRISM Genetic analyzer.

#### **4.2.10. Minipreparace DNA**

1,5 ml suspenze bakterií v LB médiu jsme centrifugovali v mikrozkumavkách (2 min, 17000 g). Pelet bakterií jsme resuspendovali v 50  $\mu$ l resuspendačního roztoku (TrisCl pH 8 25 mM, glukosa 1%, EDTA 10 mM). Následně jsme přidali za intenzivního míchání 100  $\mu$ l lyzačního roztoku (NaOH 0,2 M, SDS 2%), jemně promíchali se 150 $\mu$ l neutralizačního roztoku (3 M acetát draselný, pH 5,5) a nechali 5 minut při pokojové teplotě. Do lyzátu bakterií jsme přidali 250  $\mu$ l směsi fenolu a chloroformu (1:1) a centrifugovala (2 min, 6000 g). Následně jsme odebrali vodní fázi s DNA, kterou jsme vysráželi 250  $\mu$ l izopropanolu, a centrifugovali 20 min při 6000 g. Pelet s DNA jsme opláchli 70% ethanolem a znova centrifugovali 5 min. Usušený pelet jsme rozpustili ve 40  $\mu$ l T<sub>10</sub>E<sub>0,1</sub> s RNázou (10  $\mu$ g/ml). 3  $\mu$ l DNA jsme odebrali pro kontrolu vzorku restrikčními endonukleázami. Zbytek jsme nechali 1 hodinu ve 37°C, abyhom se zbavili RNA. Štěpené plazmidy jsme zkontovali pomocí elektroforézy. Pozitivní vzorky jsme naředili do 300  $\mu$ l vodou, přidali 10% objemu 3 M acetátu sodného a znova přečistili směsi fenolu a chloroformu. Čistou DNA jsme rozpustili v 15  $\mu$ l vody a na gelu ověřili koncentraci.

#### **4.2.11. Použitý počítačový software**

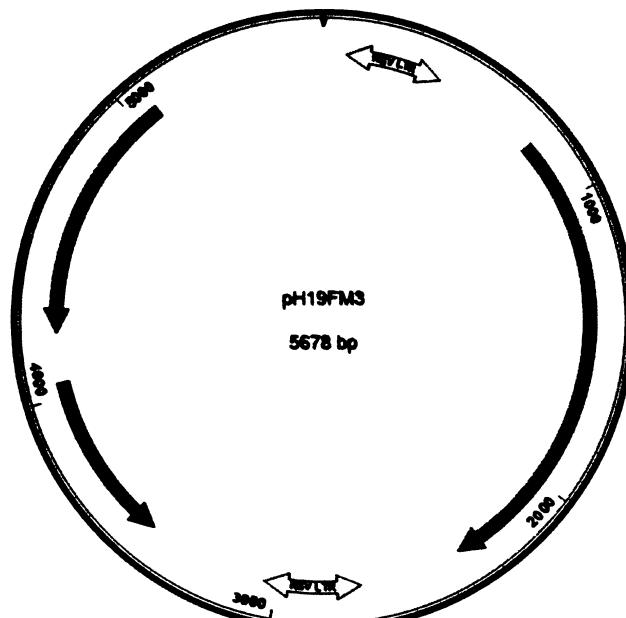
K porovnání a úpravě sekvencí jsme použili programy Chromas Lite ve verzi 2.01 (Technelysium) a Lasergene 8.0.2 (DNASTAR). Lasergene sme pak také použili k tvorbě restrikčních map konstruktů a klonovacích schémat.

Primery byly analyzovány pomocí aplikace na internetové adrese: <http://scitools.idtdna.com/scitools/Applications/OligoAnalyzer>.

Získaná integrační místa jsme ověřovali pomocí BLAT genome browser, který je k dispozici na internetové adrese: <http://genome.brc.mcw.edu/cgi-bin/hgBlat>

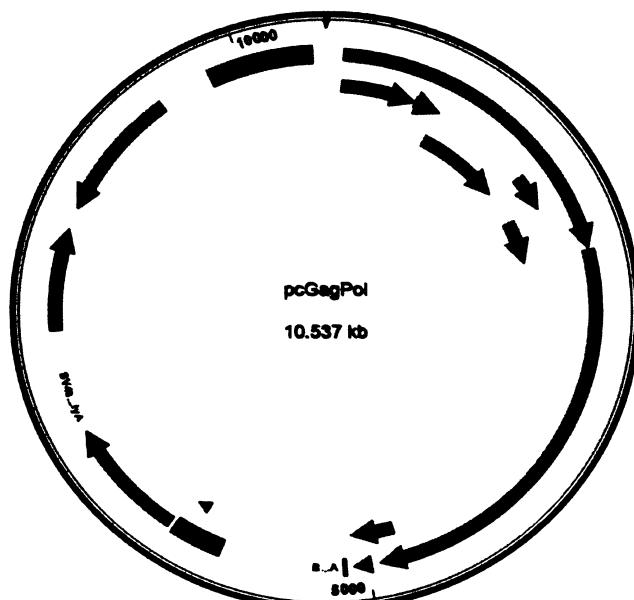
#### 4.2.12. Mapy použitých vektorů

pH19FM3



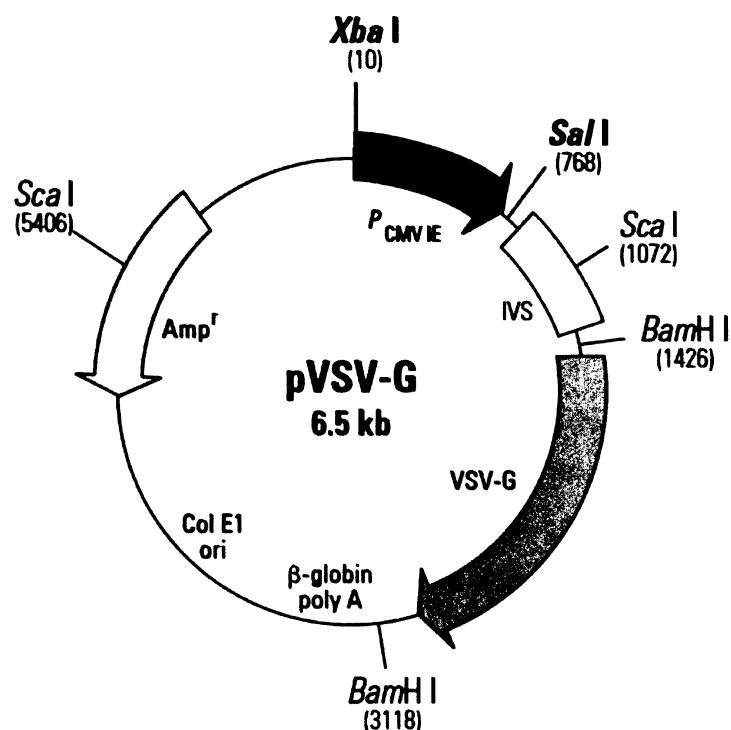
Obrázek 10 Od RSV odvozený vektor, nesoucí LTR-vsrt-LTR, byl použit pro indukci nádorů.

pcGagPol



Obrázek 11. Vektor stabilně integrovaný v buněčné linii E6.

### pVSV-G



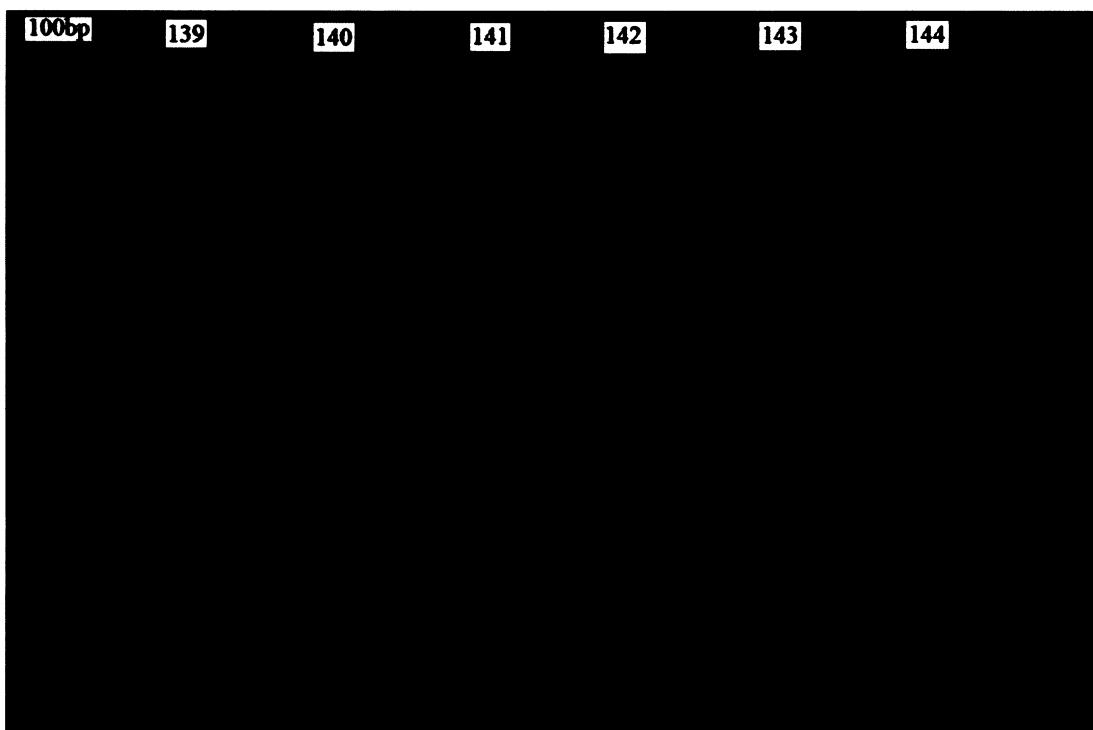
Obrázek 12. Komerčně dotupný vektor (Clontech). Použit při kotransfekci s pH19FM3

## 5. VÝSLEDKY

### 5.1. Extrakce a ověření míst integrace

Celkově jsme zpracovali 311 nádorů z 85 kuřat během tří nezávislých infekcí. Každé kuře bylo inokulováno na čtyřech místech, z čehož vyplývá incidence 91%.

Z těchto 311 nádorů jsme po I-PCR (obr. 10) a sekvenaci identifikovali 165 produktů s korektní strukturou. Ta sestává z leader sekvence s *PstI* místem na rozhraní mezi provirovou a chromozomální DNA, následuje oblast s neznámou genomickou sekvencí a konec tvoří část U3. Po odstranění provirové DNA jsme získali sekvence o délce 22 – 1741 bp. Každý takový fragment jsme prošetřili pomocí programu BLAT a sekvence s vysokou homologií (>98%) po celé délce a jednoznačným umístěním v kuřecím genomu byly označeny jako integrační místa. V případě nálezu dvou a více totožných integračních míst o stejné délce a se shodným genomickým umístěním jsme použili jen jednu sekvenci, ostatní jsme vyřadili jakožto pravděpodobný výsledek kontaminace při I-PCR, nebo koloniovém PCR.



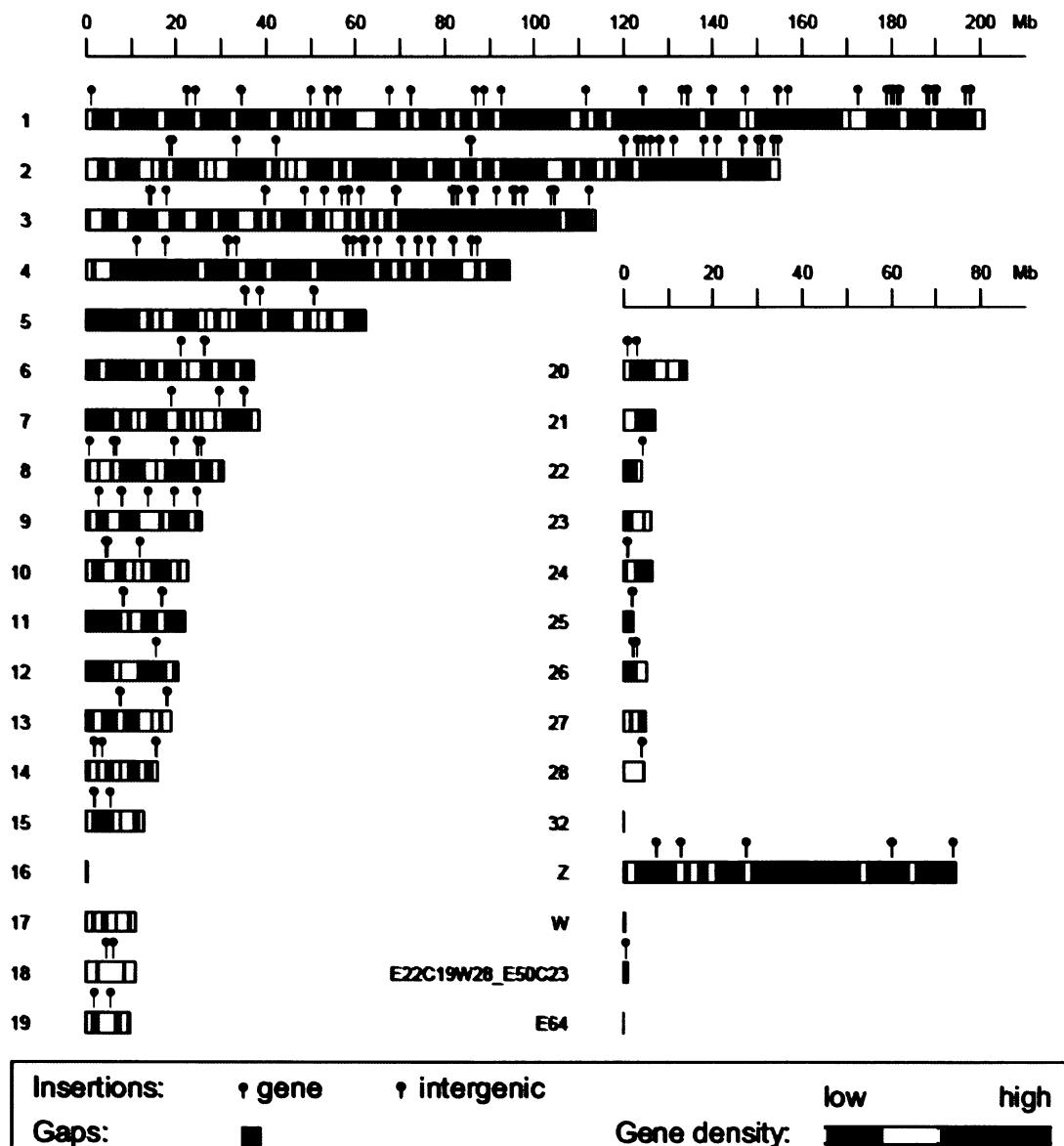
Obrázek 13. I-PCR produkty z nádorů č.139-144 na 1% agarózovém gelu.

## **5.2 Distribuce integračních míst na chromozomech**

Ze 165 fragmentů s korektní strukturou se nakonec podařilo v kuřecím genomu jednoznačně identifikovat 153 unikátní integrační místa (viz. kapitola 8. Příloha). Ostatní sekvence byly vyřazeny kvůli integraci do repetitivních sekvencí, nebo se je v současné verzi genomu nepodařilo identifikovat.

Alespoň jednu integraci ze 153 zjištěných hostil téměř každý chromozom, výjimkou byly chromozomy 16, 17, 21, 23, 27, W a E64. Ale z nich jen chromozom 17 dosahuje velikosti 10 Mb. Více než 10 integrací hostily chromozomy 1, 2, 3, 4, šlo o 38, 21, 14, 17 integrací. 5 a více integrací jsme zaznamenali v chromozomech 8, 9 a pohlavním chromozomu Z, bylo jich 6, 5, 5. Zbylé chromozomy hostily 1-3 integrace. Při porovnání velikosti a počtu integračních míst u chromozomů s více jak 25 Mb je zřetelný nadprůměrný počet integrací do chromozomů 1, 3, 4, 8 a 9. Nejnižší počet integrací vzhledem ke své délce pak hostí chromozomy 5 a 6. Pohlavní chromozom Z hostil 5 integrací, W ani jednu (obr. 14) a (tab. 5.1.).

Na chromozomu 1 je zajímavá vysoká densita integrací kolem 180 Mb a 190 Mb. V genově bohaté oblasti kolem 180 Mb došlo k pěti integracím v oblasti o velikosti 2,9 Mb a tři z pěti nastaly do genů. Oblast kolem 190 Mb je zřetelně chudší na geny, ale i tak zde došlo k k šesti integracím v sekvenci o délce 2,5 Mb, z toho tři byly do genů (obr. 15 a 16). Tyto obrázky znázorňují primární výstup z našich dat.



**Obrázek 14.** Lokalizace RSV integrací z nádorů na jednotlivých chromozomech. Integrace do genů (geny pro popsané mRNA a/nebo Unigene klastry) jsou znázorněny jako červená "lízatka", černá značí integraci do mezigenových oblastí. Barevná výplň chromozomů znázorňuje genovu densitu v rámci 1Mb velké oblasti. Černou výplní jsou vyznačeny nesekvenované části genomu.

**Tabulka 5.1.** Popisky zleva doprava: číslo sekvence; chromozom; počátek integrace; konec integrace; orientace proviru; délka sekvence; počet nukleotidů pánujících v genu; vzdálenost od předchozí integrace na chromozomu; do genu – 1, mimo gen – 0; popis genu

seq	ch	start	end	strand	seq	len	ch	span	delta_bp	genes	gene_ids
177	1	1023816	1023881	-	65	65				1	ENSGALG000000008437:1011481-1063364:KNOWN:protein_coding
32	1	22324444	22324640	+	195	196	21300628	0			ENSGALG000000008874:24199922-24224904:KNOWN_BY_-PROJECTION:protein_coding
162	1	24203959	24204415	+	456	456	1879515	1			
109	1	34416977	34417388	-	414	411	10213018	0			
160	1	34443887	34444250	+	364	363	26910	0			
59	1	49930704	49930810	-	106	106	15486817	1			ENSGALG00000023235:49925024-49931191:NOVEL:protein_coding
145	1	53774041	53774656	+	622	615	3843337	0			ENSGALG0000012655:55854406-55922770:KNOWN_BY_-PROJECTION:protein_coding
144	1	55857096	55857125	+	29	29	2083055	1			ENSGALG0000013177:67517079-67576510:KNOWN_BY_-PROJECTION:protein_coding
163	1	67523437	67523593	-	156	156	11666341	1			
119	1	72386981	72387214	-	233	233	4863544	0			ENSGALG0000015152:86600181-86644954:KNOWN_BY_-PROJECTION:protein_coding
141	1	86632024	86632112	-	89	88	14245043	1			ENSGALG0000015328:88656138-88657103:KNOWN_BY_-BY_PROJECTION:protein_coding
139	1	88656225	88656936	-	711	711	2024201	1			ENSGALG0000015403:92450603-92669952:KNOWN:protein_coding
126	1	92461776	92462673	-	895	897	3805551	1			ENSGALG0000016065:111393546-111444936:KNOWN_BY_-BY_PROJECTION:protein_coding
82	1	111406873	111407251	+	378	378	18945097	1			
155	1	124229012	124229814	-	801	802	12822139	0			
117	1	13292558	132926140	+	582	582	8696546	1			ENSGALG0000016681:132779441-132944324:KNOWN_BY_-PROJECTION:protein_coding
190	1	134132732	134133360	-	628	628	1207174	1			ENSGALG0000016716:134118363-134168963:NOVEL:protein_coding
132	1	134309167	134309227	-	60	60	176435	0			ENSGALG0000016797:139414138-139453129:KNOWN_BY_-PROJECTION:protein_coding
183	1	139436814	139437129	+	320	315	5127647	1			

seq	ch	start	end	strand	seq_len	ch	span	delta_bp	genes	gene_ids
62	1	139895920	139896380	-	460	460	459106	1	NOVEL;protein_coding	ENSGALG00000016802:139858851-139909681:
124	1	147107633	147108301	-	690	668	7211713	1	PROJECTION;protein_coding	ENSGALG00000016865:147071021-147108634:KNOWN_BY_-
164	1	154520881	154521372	+	500	491	7413248	0		ENSGALG00000016906:156784266-156890288:
41	1	156788071	156788359	+	288	288	2267190	1	KNOWN;protein_coding	
104	1	172481209	172482879	-	1730	1670	15693138	0		
101	1	179050710	179051543	-	832	833	6569501	0		ENSGALG00000017085:179733883-179752930:
140	1	179737405	179737436	+	31	31	686695	1	KNOWN;protein_coding	
143	1	180442712	180443358	+	648	646	705307	1	PROJECTION;protein_coding	ENSGALG00000017095:180442066-180460890:KNOWN_BY_-
9	1	181418325	181418677	+	352	352	975613	0		
38	1	181898772	181898820	-	48	48	480447	1	NOVEL;protein_coding	ENSGALG00000019075:181894711-181902280:
147	1	187644844	187645167	+	323	323	5746072	1	NOVEL;protein_coding	ENSGALG00000017196:187511056-187666706:
138	1	187815739	187815770	+	31	31	170895	1	KNOWN;protein_coding	ENSGALG00000017197:187746707-187981291:
121	1	188353740	188353765	-	25	25	538001	0		
18	1	189453157	189453843	+	686	686	1099417	0		
37	1	189975210	189975743	+	533	533	522053	0		ENSGALG00000017219:190082832-190092144:
17	1	190090350	190091132	+	777	782	115140	1	KNOWN_BY_PROJECTION;protein_coding	
47	1	196643098	196643559	+	461	461	6552748	0		ENSGALG00000017264:197525245-197690677:
182	1	197672017	197672439	-	427	422	1028919	1	NOVEL;protein_coding	ENSGALG00000017268:197806718-197807413:KNOWN_BY_-
2	1	197807173	197807858	+	684	685	135156	1	PROJECTION;protein_coding	ENSGALG00000007956:18449189-18682526:KNOWN_BY_-
11	2	18520005	18520126	+	121	121		1	PROJECTION;protein_coding	

seq	ch	start	end	strand	seq	len	ch	span	delta_bp	genes	gene_ids
20	2	18928436	18928515	-	79	79	408431	0			
161	2	19135030	19135327	+	297	297	206594	1	PROJECTION;protein_coding	ENSGALG00000008591:19049186-19261236:KNOWN_BY_-	
95	2	33313296	33313496	-	198	200	141178266	1	PROJECTION;protein_coding	ENSGALG00000011137:33214723-33422302:KNOWN_BY_-	
142	2	42114204	42114836	-	632	632	8800908	1	PROJECTION;protein_coding	ENSGALG0000001161341915663-42135932:KNOWN_BY_-	
165	2	85378523	85378906	-	384	383	43264319	1	PROJECTION;protein_coding	ENSGALG00000013124:85359668-85724078:KNOWN_BY_-	
192	2	85834280	85834666	-	384	386	455757	1	PROJECTION;protein_coding	ENSGALG00000013129:85793668-85836104:KNOWN:protein_coding	
180	2	119985311	119985477	+	166	166	34151031	0			
195	2	120030290	120030465	-	175	175	44979	0			
7	2	123045421	123046597	+	1198	1176	3015131	0			
48	2	124226160	124226466	+	306	306	1180739	1	PROJECTION;protein_coding	ENSGALG00000015673:124125745-124251875:KNOWN_BY_-	
111	2	125950441	125951216	+	775	775	1724281	1	PROJECTION;protein_coding	ENSGALG00000015758:125920143-125994998:KNOWN_BY_-	
40	2	128037410	128037691	+	281	281	2086969	0			
15	2	131112865	131113053	+	188	188	3075455	1	KNOWN:protein_coding	ENSGALG00000015984:131110546-131136245:	
1	2	137860023	137860324	-	310	301	6747158	1	KNOWN:protein_coding	ENSGALG00000016106:137837383-137864202:	
79	2	140884573	140885038	+	465	465	3024550	1	KNOWN:protein_coding	ENSGALG00000016122:140871172-140953283:	
74	2	146718854	146719182	-	328	328	5834281	0			
128	2	149928233	149928636	-	402	403	3209379	1	PROJECTION;protein_coding	ENSGALG00000016197:149893942-149958361:KNOWN_BY_-	
45	2	150730407	150730493	+	86	86	802174	0			
77	2	153624262	153625244	-	984	982	2893855	0			
14	2	154546783	154547141	-	359	358	922521	1	PROJECTION;protein_coding	ENSGALG00000016131:154542112-154557913:KNOWN_BY_-	

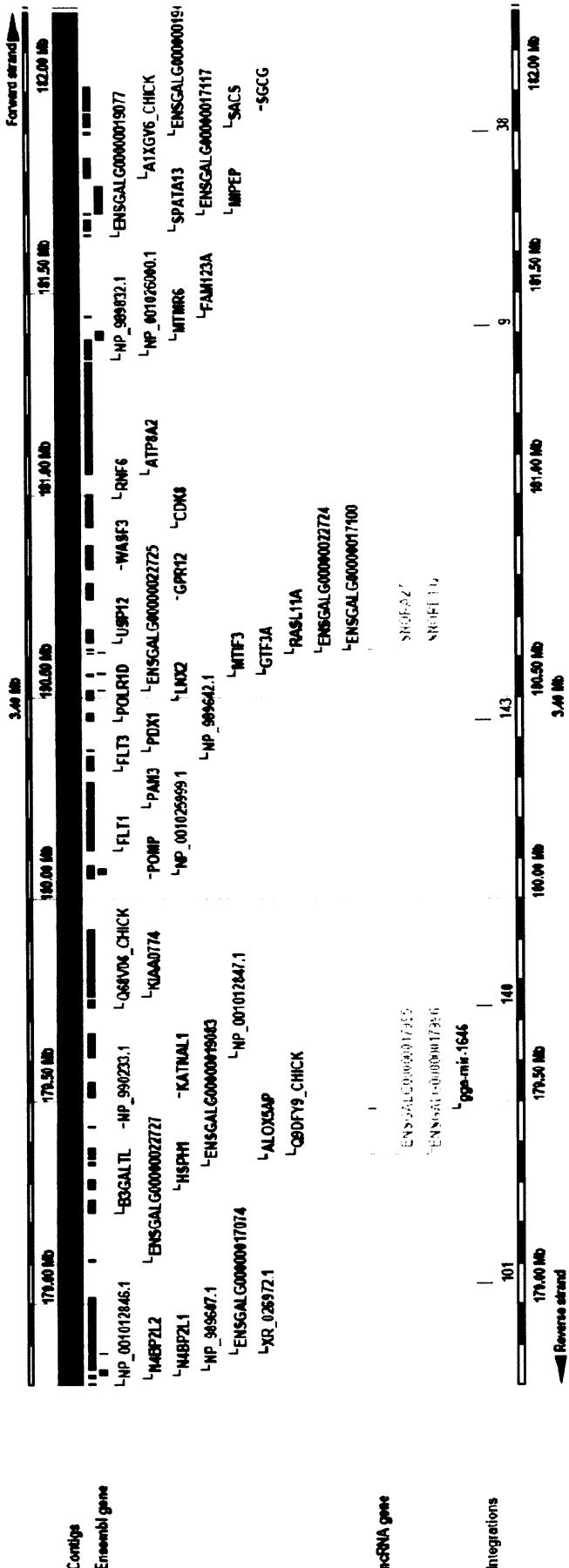
seq	ch	start	end	strand	seq	len	ch.	span	delta_bp	genes	gene_ids
134	3	14108327	14108556	+	229	229			0		
137	3	14410205	14410283	-	78	78	301878	1	ENSGALG00000009020:14397681-14432184:NOVEL:protein_coding		
71	3	17654447	17654525	-	78	78	3244242	1	ENSGALG00000013224:17638354-17695633:KNOWN_BY_-PROJECTION:protein_coding		
116	3	39889707	39889846	-	139	139	22235260	0	ENSGALG00000012297:48183036-48711189:KNOWN_BY_-PROJECTION:protein_coding ENSGALG00000012289:48466775-48513672:KNOWN_BY_-PROJECTION:protein_coding		
42	3	48474077	48474443	+	367	366	8584370	2	ENSGALG00000023054:52949913-52996184:NOVEL:protein_coding		
29	3	52958199	52958394	-	195	195	4484122	1	ENSGALG00000013892:56930192-56985664:KNOWN_BY_-PROJECTION:protein_coding		
154	3	56963631	56963955	+	324	324	4005432	1	ENSGALG00000014783:60905425-61220608:KNOWN_BY_-PROJECTION:protein_coding		
148	3	58398718	58399145	-	430	427	1435087	0	ENSGALG00000015033:68760388-68828037:NOVEL:protein_coding		
199	3	61028576	61029038	-	463	462	2629858	1	ENSGALG00000014783:60905425-61220608:KNOWN_BY_-PROJECTION:protein_coding		
122	3	68766921	68766977	+	56	56	7738345	1	ENSGALG00000015033:68760388-68828037:NOVEL:protein_coding		
184	3	69148164	69148478	+	314	314	381243	0			
191	3	81535485	81536076	+	591	591	12387321	0	ENSGALG00000015866:81766268-81876681:KNOWN_BY:protein_coding		
81	3	81858798	81858930	+	132	132	3233313	1	ENSGALG00000015874:81918143-81947237:KNOWN_BY_-PROJECTION:protein_coding		
178	3	81930324	81930413	+	89	89	71526	1	ENSGALG00000016155:85661761-85840363:KNOWN_BY_-PROJECTION:protein_coding		
58	3	82927295	82927397	-	102	102	996971	0			
22	3	85766208	85766728	+	521	520	2838913	1	ENSGALG00000016155:85661761-85840363:KNOWN_BY_-PROJECTION:protein_coding		
131	3	86475117	86475742	-	628	625	708909	0	ENSGALG00000016332:91368459-91490553:NOVEL:protein_coding		
91	3	91400520	91400921	-	401	401	4925403	1	ENSGALG00000016365:94980647-94985338:KNOWN_BY:protein_coding		
152	3	94984210	94984322	+	112	112	3583690	1			
120	3	95665868	95665890	-	22	22	681658	0			
83	3	97549142	97549492	+	350	350	1883274	0	ENSGALG00000016471:103457412:KNOWN_BY_-PROJECTION:protein_coding		
151	3	103454280	103455246	+	964	966	5905138	1			

seq	ch	start	end	strand	seq_len	ch_span	delta_bp	genes	gene_ids
135	3	104492789	104492977	-	188	188	1038509	0	ENSGALG00000016705:111880292-112142933:1 KNOWNN:protein_coding
27	3	112117336	112117542	+	206	206	7624547	1	ENSGALG00000007443:11092947-11110811:KNOWNN:protein_coding
123	4	11109631	11109952	-	321	321		1	ENSGALG00000009040:17517668-17541997:KNOWNN_BY_- PROJECTIONN:protein_coding
28	4	17533263	17533349	-	86	86	6423632	1	ENSGALG00000009880:31215264-31489638:KNOWNN_BY_- PROJECTIONN:protein_coding
170	4	31216565	31216956	+	391	391	13683302	1	ENSGALG00000010031:33115090-33243623:KNOWNN_BY_- PROJECTIONN:protein_coding
56	4	31631659	31631937	-	278	278	415094	0	ENSGALG00000010031:33115090-33243623:KNOWNN_BY_- PROJECTIONN:protein_coding
86	4	33231939	33232755	+	795	816	1600280	1	ENSGALG00000012108:59395582-59429439: KNOWNN_BY_PROJECTIONN:protein_coding
193	4	58049613	58049988	+	381	375	24817674	0	
12	4	58097682	58097970	+	284	288	48069	0	
39	4	59421853	59422241	-	384	388	1324171	1	ENSGALG00000012239:61512267-61536253:KNOWNN:protein_coding
130	4	61513271	61513411	+	140	140	2091418	1	ENSGALG00000013615:64791323-64840800:KNOWNN_BY_- PROJECTIONN:protein_coding
92	4	62140550	62142293	+	1741	1743	627279	0	
108	4	64825208	64825249	+	41	41	2684658	1	ENSGALG00000013615:64791323-64840800:KNOWNN_BY_- PROJECTIONN:protein_coding
16	4	70174773	70175332	+	562	559	5349565	0	
55	4	73976945	73980020	-	1300	3075	3802172	0	
36	4	76929880	76930117	+	237	237	2952935	0	
188	4	811757396	811757873	-	477	477	48227516	0	
93	4	85830104	85830757	+	653	653	4072708	0	
198	4	87055057	87056050	-	1004	993	1224953	1	ENSGALG00000015713:86995649-87062906:KNOWNN:protein_coding
49	5	35329453	35330580	+	1091	1127		0	ENSGALG00000010054:38502094-38538946:KNOWNN_BY_- PROJECTIONN:protein_coding
118	5	38512686	38512876	+	190	190	3183233	1	ENSGALG0000001789:20902797-20941636:KNOWNN_BY_- PROJECTIONN:protein_coding
186	5	50704620	50705203	-	583	583	12191934	0	
113	6	20939435	20939496	+	61	61		1	

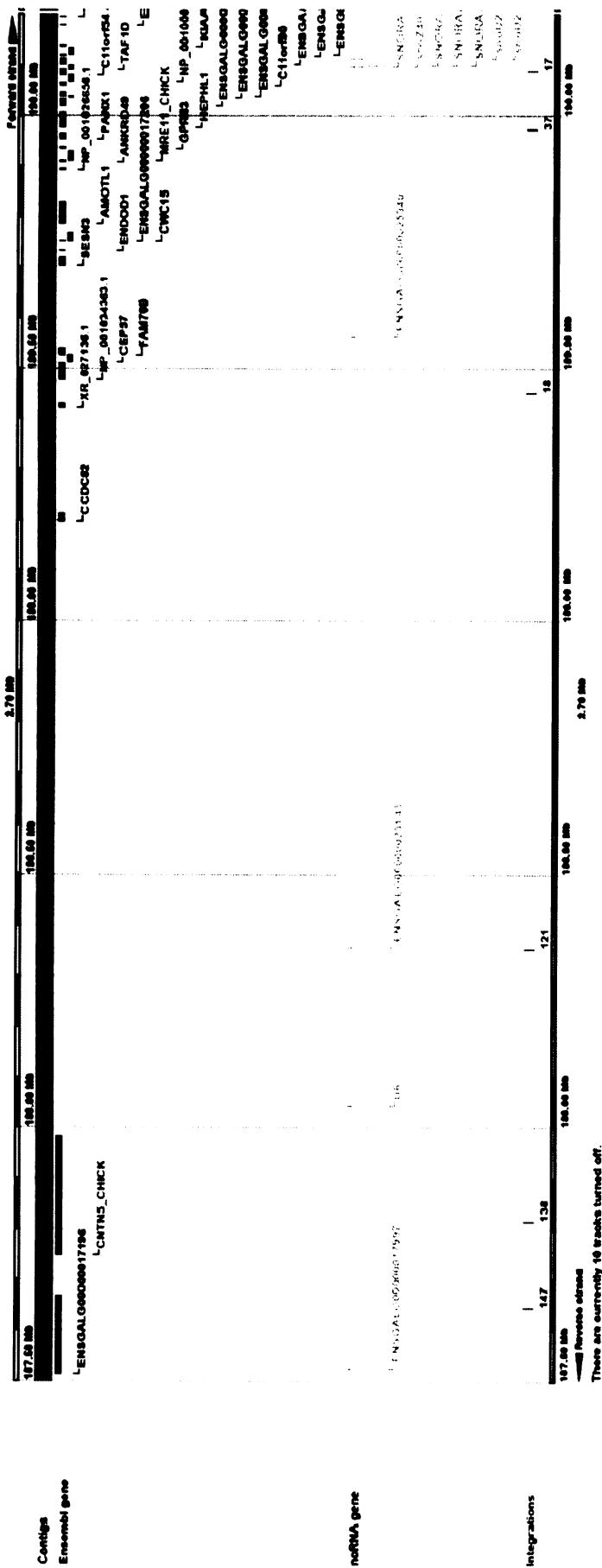
seq	ch	start	end	strand	seq	len	ch	span	delta_bp	genes	gene_ids
201	6	26415819	26416360	+	539	541	5476384		0	ENSGALG00000009357:18753689-18933272:KNOWN_BY_- PROJECTION:protein_coding	
149	7	18776651	18776746	-	99	95			1	ENSGALG00000011744:29152469-29447748:KNOWN_BY_- PROJECTION:protein_coding	
176	7	29388304	29388571	-	267	267	10611653		1	ENSGALG00000021243:455926-613642:NOVEL:protein_coding	
65	7	35083937	35084843	-	906	906	5695633		0	ENSGALG000000042056470468-6492634:KNOWN_BY_- PROJECTION:protein_coding	
200	8	496991	497889	-	900	898			1	ENSGALG00000002716:5760583-5784139:KNOWN_BY_- PROJECTION:protein_coding	
125	8	5762173	5762306	-	133	133	5265182		1	ENSGALG000000042056470468-6492634:KNOWN_BY_- PROJECTION:protein_coding	
96	8	6472650	6472837	-	187	187	710477		1	ENSGALG00000002155:2514061-2680818:KNOWNprotein_coding PROJECTION:protein_coding	
5	8	19349811	19350532	-	721	721	12877161		1	ENSGALG00000008900:19334558-19377182:KNOWNprotein_coding PROJECTION:protein_coding	
24	8	24847067	24847831	-	764	764	5497256		0	ENSGALG00000010673:25336957-25379533:KNOWN_BY_- PROJECTION:protein_coding	
158	8	25358597	25358937	-	340	340	511530		1	ENSGALG00000002155:2514061-2680818:KNOWNprotein_coding PROJECTION:protein_coding	
173	9	2587385	2587600	-	215	215			1	ENSGALG00000002155:2514061-2680818:KNOWNprotein_coding PROJECTION:protein_coding	
127	9	7861426	7863587	-	1621	2161	5274041		0	ENSGALG0000006834:13520652-13536937:KNOWN_BY_- PROJECTION:protein_coding	
175	9	13521706	13522062	-	356	356	5660280		1	ENSGALG00000013837:19274096-19387696:KNOWNprotein_coding PROJECTION:protein_coding	
150	9	19360223	19360336	-	113	113	5838517		1	ENSGALG0000009699:24336678-24348375:KNOWN_BY_- PROJECTION:protein_coding	
88	9	24346151	24346252	-	102	101	4985928		1	ENSGALG0000002774:4049737-4071553:KNOWN_BY_- PROJECTION:protein_coding	
196	10	4050784	4051283	-	499	499			1	ENSGALG0000002774:4049737-4071553:KNOWN_BY_- PROJECTION:protein_coding	
19	10	4773619	4773843	+	233	224	722835		0	ENSGALG0000004960:11673560-11824138:KNOWNprotein_coding PROJECTION:protein_coding	
43	10	11729016	11729115	+	99	99	6955397		1	ENSGALG0000007670:15273307-15495851:KNOWN_BY_- PROJECTION:protein_coding	
100	11	8350996	8351674	+	678	678			0		
50	11	1699886	16999234	+	348	348	8647890		0		
110	12	15404485	15404564	+	28	28			1		
85	13	7592753	75933322	-	590	569			0		

seq	ch	start	end	strand	seq	len	ch	span	delta_bp	genes	gene_ids
57	13	18102677	18103524	+	847	847	10509924	0			
66	14	1833008	1833715	+	707	707	707	0		ENSGALG0000004368:3390192-3424515:KNOWN_BY_- PROJECTION:protein_coding	
156	14	3402433	3402765	+	340	332	1569425	1			
78	14	15584396	15584629	+	233	233	12181963	0			
4	15	1791619	1791687	+	68	68	68	0		ENSGALG0000003413-5106007-5153114:KNOWN_BY_- PROJECTION:protein_coding	
174	15	5136147	5136633	-	486	486	3344528	1		ENSGALG0000001805-4251719-4271786:NOVEL:protein_coding ENSGALG00000003033:5851657-5893463:KNOWN_BY_- PROJECTION:protein_coding	
172	18	4262425	4262785	-	361	360	360	1			
112	18	5884405	5884687	+	282	282	1621980	1		ENSGALG0000001153-1304448-1975038:KNOWN_BY_- PROJECTION:protein_coding	
187	19	1665319	1666006	-	687	687		1		ENSGALG0000002783-5278829-5302444:KNOWN_BY_- PROJECTION:protein_coding	
171	19	5298495	5298573	-	78	78	3633176	1			
26	20	526080	526224	-	145	144		1		ENSGALG0000001050-443132-530154:KNOWN:protein_coding ENSGALG00000001420-730980-787784:KNOWN:protein_coding	
63	20	742705	742832	-	127	127	216625	1		ENSGALG00000003362-26332732-26633681:KNOWN_BY_- PROJECTION:protein_coding	
10	20	2655635	2656339	-	704	704	1912930	1			
157	22	3917092	3917343	-	136	251		1		ENSGALG00000014155-3905594-3920581:KNOWN:protein_coding PROJECTION:protein_coding	
166	24	815331	815710	-	379	379	379	0			
181	25	1542717	1542795	-	78	78		1		ENSGALG00000019796-1535163-1542999:KNOWN_BY_- PROJECTION:protein_coding	
8	25	1866088	1866211	+	123	123	323371	1		ENSGALG0000000809-1857856-1875203:KNOWN_BY_- PROJECTION:protein_coding	
31	26	2110710	2110904	+	194	194		0		ENSGALG0000001264-2573155-2661387:KNOWN_BY_- PROJECTION:protein_coding	
106	26	2609192	2609923	+	730	731	498482	1			
80	28	4115637	4115756	-	119	119		0			
84	27	7245687	7245850	-	163	163		0			
129	27	12735504	12735712	-	208	208	5489817	0			

seq	ch	start	end	strand	seq_len	ch_span	delta_bp	genes	gene_ids
179	Z	27455163	27455281	-	118	118	14719659	0	
61	Z	60029166	60029845	+	673	679	32574003	0	
185	Z	60186353	60186722	+	369	369	157187	0	
197	Z	73821888	73822891	-	1003	1003	13635535	0	ENSGALG0000006169:222577-242140:KNOWN_BY_-
136	E22	230791	230892	-	101	101		1	PROJECTION;protein_coding



**Obrázek 15.** Oblast s vysokou hustotou integrací v distální části chromozomu I. Místo integrace je označeno zelenou šípkou s číslem nádoru.



Obrázek 16. Oblast s vysokou hustotou integrací v distální části chromozomu 1. Místo integrace je označeno zelenou šípkou s číslem nádoru.

### **5.3. Preference RSV pro geny**

Pro porovnání počtu integračních míst v genech, nebo mezigenových oblastech jsme museli specifikovat integraci do genu. Tu jsme definovali jako integraci uvnitř genů pro známé mRNA a/nebo do Unigene sekvencí. Takovýchto integrací jsme získali 88, což při celkovém počtu 153 integrace znamená, že 57,5% integrací nastalo v genech. Porovnáme-li procento intragenových integrací s procentuálním obsahem genů pro známou mRNA a Unigene sekvence v kuřecím genomu, jenž činí 22,7% (REINISOVÁ et al. 2008), vyjde nám více než dvojnásobná preference pro geny v porovnání s náhodnou integrací.

Z celkového počtu 153 integrací bylo shodně s kuřecí DNA orientováno 73 vektorů, v opačné orientaci jsme detekovali 80 integrací, z toho v + orientaci proběhla integrace do genů v 39 případech (53,4%), v – orientaci ve 49 případech (61,3%).

## 6. DISKUSE

Získání integračních míst RSV v nádorových buňkách nám mělo umožnit charakterizovat místa, kde dochází k dlouhodobé a dostatečně silné expresi onkogenu *v-src*. Vliv této specifické selekce nebyl dosud studován.

Jelikož byla tato studie časově velmi náročná (zdržení zapříčiněné semirezistencí kuřat, práce s kuřaty a indukce nádorů, zpracování všech 311 nádorů) jsou zde publikované výsledky zatím pouze primární hrubá data, která ještě budou podrobena dalším analýzám. Zaměřit se hodláme zejména na data, která byla analyzována v ostatních publikacích, zabývajících se ASLV integrací v neselektovaných podmínkách (NAREZKINA et al. (2004), MITCHELL et al. (2004), BARR et al. (2005), REINIŠOVÁ et al. (2008)). Zajímat nás bude hlavně obsah GC v okolí místa integrace, vzdálenost od transkripčního počátku, přítomnost CpG ostrovů, přesná lokalizace v rámci zasaženého genu, integrace do intronu/exonu, hladina exprese zasažených genů a genů v nejbližším okolí integrace a jejich expresní charakteristika v rozdílných tkáních. Zajímavé bude také porovnaní integrací do udržovacích (housekeeping) genů, které byly včetně svého 100kb okolí v práci Renišové et al. mnohem častěji preferovány pro integraci, než tkáňově specifické geny.

Už z doposavad provedených analýz je ale patrný výrazný posun v četnosti integrace do genů, resp. transkripčních jednotek. V předchozích publikacích na buněčných kulturách Narezkina et al. (v lidských Hela buňkách), Barr et al. a Reinišová et al. (v CEF) pozorovali integraci do transkripčních jednotek ve 42%, 37,1% a 36,8%.

My jsme ale z transkripčních jednotek získali 57,5%, což značí jasně zvýšenou preferenci pro oblasti genů v integračních místech získaných z RSV-nádorů.

Tento výsledek jsme do jisté míry očekávali a připisujeme jej působení selekčního tlaku v nádorech. Ten je dán nutností dlouhodobé a dostatečně silné exprese onkogenu *v-src* pro indukci a progresi nádoru, která je ale těžko dosažitelná v oblastech s nízkou genovou densitou (genové pouště, gene deserts) nebo v heterochromatinových oblastech, zejména konstitutivních (LEWINSKI et al. 2005). Naopak v oblastech genů může být tato exprese zvyšována již přítomnými

transkripcí aktivujícími faktory, jako je přítomnost transcription factories, nebo acetylace histonů (SCHNEIDER a GROSSCHEDL 2007). Potvrzení tohoto předpokladu pak ukazuje, že jsme nejspíše zvolili správnou cestu a pomocí dalších analýz se nám snad podaří přispět k nalezení genomických oblastí umožňujících kontinuální expresi integrovaného retrovíru.

Zajímavé výsledky nám může také poskytnout porovnání počtu integrací do genů v souvislosti s orientací genu. Zde se nabízí předpoklad pro mírné zvýšení poměru ve prospěch integrací v – (antisense) orientaci, kvůli možné destrukci genomových transkriptů přispěním provirových polyA signálů, nebo stop kodonů. Zatím ale známe pouze orientaci provirových DNA vůči chromozomu, což nám nestačí, jelikož geny mohou být také v – orientaci.

Nevýhodou našeho přístupu je nemožnost zachytit všechny zaintegrované proviry, což je dáno omezenými možnostmi námi použitá restrikční metody a tím, že některé nádory mohou být polyklonální. To ale vede spíše k podhodnocení našich dat, jelikož z takovéhoto nádoru můžeme získat provirus, který nebyl příčinou onkogeneze. Jistým východiskem by mohla být analýza, ze které by byly vyloučeny nádory s více pozitivními i-PCR produkty, nebo alespoň ty, u kterých bylo nalezeno více integrací.

## 7. SHRNUTÍ

Cílem této práce bylo klonování a mapování integračních míst RSV z kuřecích sarkomů a jejich analýza, která by umožnila porovnání s údaji z ostatních publikací ve kterých byla tato charakteristika provedena bez vlivu nádorové selekce na transkripční aktivitu integrovaného proviru.

Doposavad se nám podařilo:

*Získat celkem 165 fragmentů s korektní provirovou strukturou a z nich jsme jednoznačně identifikovali 153 míst v kuřecím genomu.*

Nyní máme k dispozici předběžnou analýzu primárních dat, ze které vyplynulo:

*Vzorek 153 unikátních sekvencí vykazuje o 20% větší preference pro oblasti transkripčních jednotek, než bylo pozorováno v neselektovaných buněčných kulturách.*

Z toho lze usuzovat:

*Námi získaná integrační místa nejspíše skutečně reprezentují genomické oblasti, které umožňují trvalou a dostatečně vysokou expresi viru, a tudíž je náš model vhodným nástrojem k jejich vyhledávání a studiu.*

Práce bude pokračovat kompletní celogenomovou analýzou získaných dat, ve které se zaměříme na GC oblasti, CpG ostrovy, transkripční starty, integrace do intronu/exonu, orientaci integrace vůči zasaženému genu a na hladinu exprese těchto genů.

## 8. SEZNAM LITERATURY

- Barnard R.J., Elleder D., Young J.A. (2006): Avian sarcoma and leukosis virus-receptor interactions: from classical genetics to novel insights into virus-cell membrane fusion. *Virology*. Jan 5;344(1):25-9.
- Barr S.D., Leipzig J., Shinn P., Ecker J.R., Bushman F.D. (2005): Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. *J Virol*; 79: 12035-12044
- Boeke, J.D. and Devine, S.E. (1998): Yeast retrotransposons: finding a nice quiet neighborhood. *Cell* 93, 1087–1089
- Bor Y.C., Miller M.D., Bushman F.D., Orgel L.E.(1996): Target-sequence preferences of HIV-1 integration complexes in vitro. *Virology*. 1996 Aug 1;222(1):283-8.
- Botbol Y., Raghavendra N.K., Rahman S., Engelman A., Lavigne M. (2008): Chromatinized templates reveal the requirement for the LEDGF/p75 PWPP domain during HIV-1 integration in vitro. *Nucleic Acids Research*, 2008, Vol. 36, No. 4 1237-1246
- Bowen NJ, Jordan IK, Epstein JA, Wood V, Levin HL. (2003): Retrotransposons and their recognition of pol II promoters: a comprehensive survey of the transposable elements from the complete genome sequence of *Schizosaccharomyces pombe*. *Genome Res*. Sep;13(9):1984-97.
- Brady T.L., Fuerst P.G., Dick R.A., Schmidt C., Voytas D.F. (2008): Retrotransposon target site selection by imitation of a cellular protein. *Mol Cell Biol*. Feb;28(4):1230-9.
- Bushman F.D. (2003): Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. *Cell* 115, 135–138 .
- Bushman F.D., Lewinski M., Ciuffi A., Barr S., Leipzig J., Hannenhalli S., Hoffmann Ch. (2005): Genome wide analysis of retroviral DNA integration. *Nat. Rev. Microbiol.* 3, 848–858
- Carteau S, Gorelick R.J., Bushman F.D. (1999): Coupled integration of human immunodeficiency virus cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. *J. Virol.* 73, 6670–667
- Cereseto, A. et al. (2005) Acetylation of HIV-1 integrase by p300 regulates viral integration. *EMBO J* 24, 3070-3081
- Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, et al. (2005): A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* 11: 1287–1289.
- Ciuffi A., Llano M., Poeschla E., Hoffmann C., Leipzig J., Shinn P., Ecker J.R., Bushman F. (2005): A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.* 11, 1287–1289
- Crise, B., Y. Li, C. Yuan, D. R. Morcock, D. Whitby, D. J. Munroe, L. O. Arthur, and X. Wu. (2005): Simian immunodeficiency virus integration preference is similar to that of human immunodeficiency virus type 1. *J. Virol.* 79:12199–12204

Delelis O., Parissi V., Leh H., Mbemba G., Petit C. (2007): Efficient and Specific Internal Cleavage of a Retroviral Palindromic DNA Sequence by Tetrameric HIV-1 Integrase. PLoS ONE 2(7): e608. doi:10.1371/journal.pone.0000608

Derse D., Crise B., Li Y., Princler G., Lum N., Stewart C., McGrath C.F., Hughes S.H., Munroe D.J., Wu X. (2007): Human T-cell leukemia virus type 1 integration target sites in the human genome: comparison with those of other retroviruses. J Virol. Jun;81(12):6731-41.

Doi K., Wu X., Taniguchi Y., Yasunaga J., Satou Y., Okayama A., Nosaka K., a Matsuoka M. (2005): Preferential selection of human T-cell leukemia virus type I provirus integration sites in leukemic versus carrier states. Blood 106:1048-1053.

Elleder D., Molder D.K., Trejbalova K., Svoboda J., a Federspiel M.J. (2004): Two different molecular defects in the Tva receptor gene explain the resistance of two tvar lines of chickens to infection by subgroup A avian sarcoma and leukosis viruses. J Virol. Dec;78(24):13489-500.

Elleder D., Pavlícek A., Paces J., Hejnář J. (2002): Preferential integration of human immunodeficiency virus type 1 into genes, cytogenetic R bands and GC-rich DNA regions: insight from the human genome sequence. FEBS Lett. 2002 Apr 24;517(1-3):285-6.

Engelman Alan (2005): The ups and downs of gene expression and retroviral DNA integration. PNAS February 1, 2005 vol. 102 no. 5 1275-1276

Farnet C.M., Haseltine W.A. (1991): Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. J Virol 65: 1910-1915.

Faschinger A., Rouault F., Sollner J., Lukas A., Salmons B., Günzburg W.H., Indik S. (2008): Mouse mammary tumor virus integration site selection in human and mouse genome. J Virol. 2008 Feb;82(3):1360-7.

Faure A., Calmels Ch., Desjobert C., Castroviejo M., Caumont-Sarcos A., Tarrago- Litvak L., Litvak S. a Parissi V. (2005): HIV-1 integrase crosslinked oligomers are active in vitro Nucleic Acids Research, 2005, Vol. 33, No. 3 977-986

Ferber M.J., Montoya D.P., Yu C., Aderca I., McGee A., Thorland E.C., Nagorney D.M., Gostout B.S., Burgart L.J., Boix L., Bruix J., McMahon B.J., Cheung T.H., Chung T.K., Wong Y.F., Smith D.I., Roberts L.R. (2003): Integrations of the hepatitis B virus (HBV) and human papillomavirus (HPV) into the human telomerase reverse transcriptase (hTERT) gene in liver and cervical cancers. Oncogene 22, 3813-3820

Gudkov A.V., Obukh I.B., Serov S.M., Naroditsky B.S. (1981): Variety of endogenous proviruses in the genomes of chickens of different breeds. J Gen Virol 57: 85-94.

Hacein-Bey-Abina, S. et al. (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCIDX1. Science 302, 400-401.

Hackett C.S., Geurts A.M., Hackett P.B. (2007): Predicting preferential DNA vector insertion sites: implications for functional genomics and gene therapy. Genome Biol.;8 Suppl 1:S12.

Havecker, E. R., X. Gao, and D. F. Voytas. (2004): The diversity of LTR retrotransposons. Genome Biol. 5:225.

Hejnář J., Elleder D., Hájková P., Walter J., Blazková J., Svoboda J. (2003): Demethylation of host-cell DNA at the site of avian retrovirus integration. Biochem Biophys Res Commun. Nov 21;311(3):641-8

- Katz R.A., Merkel G., Kulkosky J., Leis J., Skalka A.M. (1990): The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro. *Cell* 63: 87–95.
- Kirchner J., Connolly C. M. a Sandmeyer S. B. (1995): Requirement of RNA polymerase III transcription factors for invitro position-specific integration of a retroviruslike element. *Science* 267(5203): 1488–1491
- Langley R.J., Hirsch V.M., O'Brien S.J., Adger-Johnson D., Goeken R.M., Olmsted R.A. (1994): Nucleotide sequence analysis of puma lentivirus (PLV-14): genomic organization and relationship to other lentiviruses. *Virology*, Aug 1;202(2):853-64.
- Lee, M.S. and Craigie, R. (1998) A previously unidentified host protein protects retroviral DNA from autointegration. *Proc. Natl. Acad. Sci. USA* 95, 1528–1533
- Levin A., Hayouka Z., Helfer M., Brack-Werner R., Friedler A., Loyter A. (2009): Peptides Derived from HIV-1 Integrase that Bind Rev Stimulate Viral Genome Integration. *PLoS ONE* 4(1): e4155.doi:10.1371/journal.pone.0004155
- Lewinski M.K., Bisgrove D., Shinn P., Chen H., Hoffmann C., Hannenhalli S., Verdin E., Berry C.C., Ecker J.R., Bushman F.D. (2005): Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J Virol*, 79:6610-6619
- Lewinski, M.K. et al. (2006) Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog* 2, e60
- Li M., Mizuuchi M., Burke T.R. Jr and Craigie R (2006): Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J*. Mar 22;25(6):1295-304.
- MacNeil A., Sankalé J.L., Meloni S.T., Sarr A.D., Mboup S., Kanki P. (2006): Genomic Sites of Human Immunodeficiency Virus Type 2 (HIV-2) Integration: Similarities to HIV-1 In Vitro and Possible Differences In Vivo. *J Virol*. Aug;80(15):7316-21.
- Manchado-Corvo P, Martín-Duque (2006): Viral gene therapy. *P.Clin Transl Oncol*. Dec;8(12):858-67.
- Maxfield L.F, Fraize C.D., Coffin J.M. (2005): Relationship between retroviral DNA-integration-site selection and hostcell transcription. *Proc. Natl. Acad. Sci. USA* 102, 1436–1441
- Mi S., Lee X., Li X.(2000): Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403 (6771): 785–9.
- Miller M.D., Farnet C.M., Bushman F.D. (1997): Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol* 71: 5382–5390
- Mitchell R.S., Beitzel B.F., Schroder A.R., Shinn P., Chen H., Berry C.C., Ecker J.R., Bushman F.D. (2004): Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol*. 2004 Aug;2(8):E234.
- Narezkina A., Taganov K.D., Litwin S., Stoyanova R., Hayashi J., Seeger C., Skalka A.M., Katz R.A. (2004): Genome-wide analyses of avian sarcoma virus integration sites. *J. Virol*. 78, 11656–11663
- Nowrouzi A., Dittrich M., Klanke C., Heinlein M., Rammling M., Dandekar T., von Kalle C., Rethwilm A. (2006): Genome-wide mapping of foamy virus vector integrations into a human cell line. *J Gen Virol*. May;87(Pt 5):1339-47.

Nowrouzi, A., M. Dittrich, C. Klanke, M. Heinkelein, M. Rammling, T. Dandekar, C. von Kalle, and A. Rethwilm (2006): Genome-wide mapping of foamy virus vector integrations into a human cell line. *J. Gen. Virol.* 87: 1339–1347

Odawara T., Oshima M., Doi K., Iwamoto A., Yoshikura H. (1998): Threshold number of provirus copies required per cell for efficient virus production and interference in moloney murine leukemia virus-infected NIH 3T3 cells. *J Virol* 72: 5414–5424.

Ozawa T, Itoyama T, Sadamori N, Yamada Y, Hata T, Tomonaga M, Isobe M. (2004): Rapid isolation of viral integration site reveals frequent integration of HTLV-1 into expressed loci. *J Hum Genet.* 49(3):154-65.

Pannell D, Osborne CS, Yao S, et al. (2000) Retrovirus vector silencing is de novo methylase independent and marked by a repressive histone code. *Embo J*; 19: 5884–5894.

Proc. Nati. Acad. Sci. USA Vol. 91, pp. 5913-5917, June 1994 Biochemistry

Pruss D., Bushman F.D., Wolffe A.P. (1994a): Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core.

Pruss D., Reeves R., Bushman F.D. a Wolffe A.P. (1994b): The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *J. Biol. Chem.* 1994 Oct 7;269(40):25031-41.

Pryciak P.M., Varmus H.E. (1992): Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* 1992 May 29;69(5):769-80.

Reinišová M., Pavlíček A., Divina P., Geryk J., Plachý J. a Hejnar J (2008): Target Site Preferences of Subgroup C Rous Sarcoma Virus Integration into the Chicken DNA. *The Open Genomics Journal*, 2008, 1, 6-12

Roe, T. et al. (1993): Integration of murine leukemievirus DNA depends on mitosis. *EMBO J* 12, 2099-2108

Rosenbluh J., Hayouka Z., Loya S., Levin A., Armon-Omer A. (2007): Interaction between HIV-1 Rev and integrase proteins: a basis for the development of anti-HIV peptides. *J Biol Chem* 282: 15743–15753

Sandmeyer S. (2003): Integration by design. *Proc. Natl. Acad. Sci. U.S. A.* 100, 5586–5588

Segura-Totten, M. and Wilson, K.L. (2004) BAF: roles in chromatin, nuclear structure and retrovirus integration. *TRENDS in Cell Biology* Vol.14 No.5 May

Shun M-C., Raghavendra N.K., Vandegraaff N., Daigle J.E., Hughes S., et al. (2007): LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev* 21: 1767–1778

Schaefer-Klein J., Givol I., Barsov E.V., Whitcomb J.M., VanBrocklin M., Foster D.N., Federspiel M.J., Hughes S.H. (1998): The EV-O-derived cell line DF-1 supports the efficient replication of avian leukosis-sarcoma viruses and vectors. *Virology*. Sep 1;248(2):305-11.

Schneider R. a Grosschedl R. (2007): Dynamics and interplay of nuclear architecture , genome organization, and gene expression *Genes Dev.* 21: 3027-3043

Schröder A.R., Shinn P., Chen H., Berry C., Ecker J.R., Bushman F. (2002): HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110, 521–529

Svoboda J., Machala O., Deozánek T. (1968): Rescue of Rous sarcoma virus in mixed cultures of virogenic mammalian and chicken cells, treated and untreated with Sendai virus and detected by focus assay. *J Gen Virol.* May;2(3):461-4

Tan W. et al. (2006) Human Immunodeficiency Virus Type 1 Incorporated with Fusion Proteins Consisting of Integrase and the Designed Polydactyl ZincFinger Protein E2C Can Bias Integration of Viral DNA into a Predetermined Chromosomal Region in Human Cells. *Journal of virology*, Feb. 2006, p. 1939–1948

Tan, W. et al. (2004). Fusion proteins consisting of human immunodeficiency virus type 1 integrase and the designed polydactyl zinc finger protein E2C direct integration of viral DNA into specific sites. *J. Virol.* 78:1301–1313.

Trobridge G.D., Miller D.G., Jacobs M.A., Allen J.M., Kiem H.P., Kaul R., Russell D.W. (2006): Foamy virus vector integration sites in normal human cells. *Proc Natl Acad Sci U S A.* Jan31;103(5):1498-503.

Tsuruo T., Baluda M.A. (1977): Integration of proviral DNA in chicken cells infected with Schmidt-Ruppin Rous sarcoma virus is not enhanced by DNA repair. *J Virol* 23: 533–542.

Vandegraaff N. a Engelman A. (2007): Molecular mechanisms of HIV integration and therapeutic intervention. *Expert Rev. Mol. Med.* Vol. 9, Issue 6, February 2007, DOI: 10.1017/S1462399407000257

Weidhaas J.B., Angelichio E.L., Fenner S., Coffin J.M. (2000): Relationship between retroviral DNA integration and gene expression. *J. Virol.* 2000 Sep;74(18):8382-9.

Wu X. a Burgess S.M. (2004): Integration target site selection for retroviruses and transposable elements. *CMLS, Cell. Mol. Life Sci.* 61 2588–2596

Wu X., Li Y., Crise B., Burgess S.M. (2003): Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300, 1749–1751.

Wu X., Li Y., Crise B., Burgess S.M., Munroe D.J. (2005): Weak Palindromic Consensus Sequences Are a Common Feature Found at the Integration Target Sites of Many Retroviruses. *J Virol.* 2005 Apr;79(8):5211-4.

Xie, W. et al. (2001): Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. *Mol. Cell. Biol.* 21(19): 6606–6614

Zhu Y., Dai J., Fuerst P. G. a Voytas D. F. (2003): Controlling integration specificity of yeast retrotransposon. *Proc. Natl Acad. Sci. USA* 100, 5891–5895.

## 9. PŘÍLOHA

#1

GTTGGGTAATAACTGAAAGCTGAGATTAATTATATACATTCCCTCTTACTGG  
TATGACAGATGAATTAATGTTCTGGGTTCTGTAGTCAGATTGTTATTAGAAATTGG  
AATAACACTGTGTATTATTCTGTAGGAGAACGCTGCATAAATTACTGCTGTTAGTGAA  
CCCAAGAACGTTAGAGCTCTGAATTAGTCTGCTCCTAGATTAGCTTCCTAAAAACAAGGG  
GAAAAGTCCTTATGTGAATAAGGTGATTCTGATTCTTTTTTTTTCCCC  
CCCAT

#2

GAACCCAACCATTGACCTGGATTGGGTTGCAAGTGTGAAGGTCTCCGGAAAGAGGAG  
TACACGAAGCAGAACCTAAAGAGACTCTGGTAGTGCAGATAATATGCCCATCACAG  
TTCCAGGTATTGGAAGAAATGCTAAAAATAGCTGTGAGTGATGGTTCTGCAAGAT  
TCTTCTCATCCACATACTTCAGATTGATGTCATCATGGAGATAACACAGGTTAAGTGTTCCCT  
TTAAAACACACTACAGCCTAATTGAGATAACACAGGTTAAGTGTTCCCT  
GGGAGATGATTTCTTAAATGTTACCAATAACAAACTTCAGAATATGTCTATATTCT  
AAATACAAGGAATATTGTCTAGCACCTGAATTAGGATTGGATCTGCTCATCCTCCACC  
TCTAACCTGCCAACAACTGGGAGGAAATCACTATCTGCTACTGCATTCCCAGATG  
AAAAATGGGAATGATCCTTAACCTGCGAACACTGGGAGGAAATGTTCACTTACCATG  
GCTGAAAAATGTCTGGAGATCATATGTGTTACTCCACAGCCGAGAAACATCCTGAGATC  
AGGTGTGCTGAAGGAAGCCACA

#4

GAGCTTCTGAGCAGAGGCACTGTGCTGTTGGACACACGGGAGGGCTGGCAGGAA  
CCGGTGTGA

#5

ACATTCAGTCTCTTCTATTGACTTTGAAAAGAGTCCTTATGTGTACCAAGTGAAT  
AACAGAAGTGTGAATACATCTAACAGTCCTTAGATAAACAGTATGTGTAAGGAGGCTCT  
GTTTAGATATAATCTCCCTAGTCTGTCAGCTACATATGCCAAATCGTGAGGCTGTCAT  
GGTCATTACATGCCGTTGCTGAAAAAATCCCTACCGGGAGAATTACATAATGGAAA  
ACGGAAATAGATAACAATTGCCAGCAAGAGCTCTTAATCTGCTCTTCACT  
AGCTACCAGTGTCTGATTTTATATGTGACTCCCAGATGTGTAAGTGATGTGCCTCAAA  
ATCCCACACATGGATCTGTACGTGCCAGTAAATAACCATGGTAAGAAGGTGAGGAAA  
GCATTCAAACAACAAACCGAGTTAGCTGCTCATGGATGCAATGAATTGTCATCTGAA  
CATTTCAGAGTTGATGTTGCTGCACTACTGTTGGGGTAGAGGATTAGCATCA  
TATCTGAGTCAGCAAGAGTTAAGCACCCTAGAAATTAGAAATCTACTTAGTAGTAAAT  
TAGGCCCTCTGTTGCTGAACAGTGTAAAGAAATTGTTGAGCACCCTGAGACTA  
TGGCATTGTCAGGGAGTGTGCTAATGAAAGTTAGAGCCCCATAAGCCC

#7

GCCCACGTATCTGTTACAGTATTCACTGGAAGGGAGTTCAGCCAAAGTACCTCCATGCA  
GCTCCTACTCGGGGGAGTGTTCATAGCAGATGGTAGTGGCATTAGCAAATGGAGGAC  
TAAGTGCTTCTCTTTAACAAACACCATGAGAGTGCTATCAATCAATTAAATTAAGC  
AGCAGTAACCTTGAGCTTCTTCACTTGCCTGCACTGTGGCTAAATATCTG  
TAGACCTGCTTATGTGAAAAATGAAGCCTCCTGGAACTCTAAATGCTGCCCTGATCT  
TGAGTAAGATTATGAAATTCTACTAATCCCTAAATGAACAATCAGTTAATGAGTTAATT  
TGTTAGGGGGGGCTGTGTCATGTCAAAGCGCTCAAGAGATAAGCCATAAAAAGACC  
ATGGAAGATTGAAAGTGTGAGAATGTTGATTGGAAAGAAAGACAAGGTTGAAACA  
AGGCCTCTGAAATATGTAAGTCCACCTATATCTCAGATTCACTGTTAGGAAACTGTT  
GCTGACAGTTCTCCACCTGTCCTTCATTTTAAATTCTATTGAAAGAAAAATGCC  
TACTTCTGATTAAATGCTCCCGTGTAGAAGGGAGCTAAAGGAGCTAAAGCTAAAGG  
AACTGTAACAGACCTCAGAAATACCTGAAGGTAAAGATAACAGCTTGAACATTGGAA

GCTTGTCCATCTGCAATTGGACACAATGTGCTATGTATGCCAAAACAGGCACAAATCCC  
AACATAAAACTAGAGGTGAGTGCAGTATGAAGCAAGCTGCATAAAAACTTAAGGA  
GCTTTTTCTTCTTTCTTTCTTTCTTTGAGACATGATGTCAGGGGGATG  
TCATTTCAAACAAATGAGGATAGGCAAACAAAGGAGGCTCGTTCATCTATAAAAGC  
AAAGTATTACAAATAAATCTGCCATTGTGAAATGTGGAGCTCATAATGATTTCAGT  
GTCTTCTGCCTTATAGGAACAGTTCCCTTAAGAAATTCTAGTTGAAATACTGATTGTT  
TACAGGAAAATATCTGATTACCAAAGGCTGTATCATGTACAGTTCACTCTGTGCGGTA  
GCCAATTGCTGGAACAGATAAAGGGTAGTCAGTAGTCTTATGCAAACTCC

#8

TCGGAGATAGGAGTCCTCCAGTGCACTCAGCACTGGGTGACACTGAGTCTGGTCAGGTT  
GATGGTTGGACTTAATGACTGTGAGGGCTTTCCAACCTAATGCTCTGAGCCTGACCC  
TAC

#9

GGGGGTTGGTAGGGCTCCTCCGAAGATGCCAGACCTCAACGTTGCCATTTCACGCA  
TGTGTGAATTTCACAAATGATCTGTGCTTGATGATGTTCCCTACCTGGATGTTTG  
CAGAGGCAGCGAGAAAGAAGGCTGCGAGCCTTATAACCTGATTGGTTAAACAAACT  
CAGCTAACGCTCATTTCTCTTCCAGCCTTCACTGGGTGAAAGGCTCAGCACAGCAGAG  
CATTACCACCGCCCGTACAGCTGCTTCCCTCCAGGGACCTGGTTCTGCTGGCATGA  
AGCGAACCGGAACAGAAGTGGGCAGTGGGGGATGGCTCTGCTGTGGG

#10

GTGGATGTACATGGAGTCTGGACAGAGGATGGGACCTCCTGTGCAGATCCAGTGCA  
CGCTCCTCTGCATGGAAGATGGGAGCCTCAGTTGCTCATTGGACTAGAAAACGTGATT  
ATTTGCTGTAGCTGTGGTAGTTGTACGTTCTCTCCCTCCGTTGAAAGGC  
AAAGACATGGACAATGAATCAGGCACTGAATGGGTGAGACTTGTCTGAAGTGGTCAGG  
CCTGCAAAAGGAGGTGCAAATGGGTGCTTGTCCAAGCATGGCATTGTTCATGGACA  
GCTGCATTCTGTGCTGGTGGTTCTGTGCGAGCTGTGAGGCTGCGTATTGATT  
ACAAGGGCTTGACAAAACCCCTGGCTCTCAGCAGCACCCAGGAGCTGCTTACCTTC  
CAACCCCTGTGCCAGAGCCTCTGGCTGGAGCAGCCTCTGGCTGCTGGTGTTC  
GCACCACTGGTTGATATCTGGAGTAGTACACAGTCACGAGCTCTGGCAAGTTCACAGCCC  
ATCATGTTAGCAGGTATCTGCTCCCTTTAGGGAAATCCCTCAAGGAAATGGATGACCT  
TCCTGCTGTTGGTTTGACATCCGCATACCTGAGAGTTGAGTGCATTCTGCATTGCCA  
CTGAGTATCTCCAAGGGATGTTCTGACACTGAACAAAGAGGAC

#11

TAACCTCGGTATTGCTAAATCTGGTATGAGACTGAATAGAGCTGTTGGAAATCTAAT  
TTGCACTCTATTCTGGTATACATCTCCACTGGTGATAAGATATTTCATGCTGTAC

#12

GGCTCCCCCGCAGCACGGCTCTGGCTGAATGCAAGCTGGCATCTCCGGTGCCTACG  
TTAAAGCACTGCGAGATGTTAATGTAACCTCATTCTAACACACCTCCACCGCGTGT  
GCAGAGGGATAACACAGCAAAGCTGCACGACGTGCCCGCGCAGAGCGAGCAGCCATGC  
CAGGGCCTGGCTGGAGTGTGGCCGCGTGTGTCAGCGCCTGCCTGGTGGCGGGTGTG  
GAAGCGGCCACACGTGGATGGCTTCAGTCAAGTTGATGATTTC

#14

AAGATTGCTCACAGATGCGATGAGAGCTTCACGCTGCCAAAACACTGCATTGCTGTGCTG  
AACCAACAGGCTGCTGCAAGGGGGTAAAGTTCTGCTGCACCTCTCCCTGGGGTGGCA  
CCTCCTGTATGTGCTGATCATGAGTAAAATGCTTTGGAGAAGAGATGGCTGAG  
ACTCATGCCACATTCAACCTCTCCTGAATTAACTTAGACCTGGCCCTGCATTAAT  
TGCACGTGTAACACGGAAAAAGTCACATAATAACCCCCCTGCAATTAAATTGTGCC  
TTTACTGCATAAAATCTATCATTGATACGATGAACCCGCTGTGGCTAGCAGAGC

#15

GTAGATGACTGAGTTGATTTGGATGTTGCTGAGACAGACTGATGCAGCATGCATAGAA  
GTAGTTCACTGTTGAAACAGCTTCCCTTGACTTAATTAGGCTAATAACATCAGT  
AACCACTTACATCCCTGATCTAGAATGTCATCAGGTGCAGAGCATTGTTCTGCTGAA  
GAGACAC

#16

GCCAAAACAATCTCTAGTAAAATATGAGATCTAACAAATTAGGCATAAAGAC  
AGTGGGCGAGATATAGCTGATGAAAGGAGCCACATACATCCCCAAACTGAGTCTGTG  
CTTGTGAAGGATTCACTATTGAATGCCAGTCGCAAAACACAAACACCACAGATAC  
TTTGTCTGAACCTAACCATACACCAACTGTGGGAGAGAGCTGTGTGGCACGTGGTGT  
GTCCCCACAGCTCCTAAAAAGATGTGGATCCAGAAGACTTACCTGGATTGCACTGATCCG  
CCCTTAAATCACTCTCCTCCCAGAGTAAGAAAGGCCACATGAGGCAATGGGTCTCAG  
ATGTTTGTGCTGGAGTAATTCTTTCTGTCTCTGAACAAATTATTAATCAAACCTTTGA  
CAATGCATTTCAAATAAAAACAGTCGGGGGAATTCATACATCCTGAATTAAATCA  
AGTGGTATAAAAATATAACTAA

#17

GCTGGTTAGAGCATTACATATGCCAGTCCTCAAACCTCCAACTCAGTGCAGACAAAATA  
TGTAATATGCTGAATCTAATTGAGGGACAGGTCACTTATCTGCTCTCGTAAGCCAAA  
GTCGCACTGCTTGAAGAAAGCCAGGCTTCTAACGACAGCAATTGCTCACTGTGTC  
TTGCTACCTAACGTTACTTTCCAGCATTGGTACTATGGACTTCAACCTGGCTCC  
CGATTCTTATGGACACGTGTTCCCTTCAAAGGGAGAAGCTGGTAGCGTGTCTCA  
AAGCAAAAAAAAGCCAGCAGCAAGTCTACAGCATTGGCTCAATACAGTACAGAGCC  
CTAACAAAAAGCTTGTTCAGAACAGAAATGTGGAGTTCTTGAGATGAGCTACAT  
TTCAGTCAGCCGTGCAACTTCATAATCAGCTATGAGGGCAAACCACAGCTGTGGGACA  
AGCAGCTGGTACATGCTACCTATAAACAGCCCCAACATTAACCTCTTAGCCACA  
TGCCTGCTGCACAAAACAGGCTCAGGCCTGGAGAAAAGCAAAAAACCTAACGAC  
ATTCTCTCAAAAAGAAAAAGGTCTCCTGGAGCAAGCAGTAGAGAAAGAAAG  
CTATCTGTAGAGTGTATTCCATCTTACTTGGCTTATCCACCTCATGGTGATTCTCC  
CTTCTCTGATCAAAGGAAAAGGAAACGGGCTGGATGTGGCTGTTGTGGACGGACT

#18

GCCTTCTCTCTCCAGGCTAACAGCCCCAGCTCTCCAGCCTGTCTGCAGAGGGGTG  
TTCCATCCCTGGATCATTCTGTGGCCCTCTCCGGATGCATTCAAACAGCTCCACATCT  
CTCCTGAACCTGAGGACTCCGCATCTGGACCTTCAGGTCCAGCACAGAGCGGAGGGCAG  
GATCACCTCCCTCGCCCTACTGGCCATGCTGCTCTGGCTGTAACACAGGATACAGTTGGC  
TCTCTGAGCTGTGGGACAGTGTCTGGCTCATACCCAGCTGCCATCCACCAAGTACTCC  
CAAGTCCTTTAACAGGGCTGTCACTCCTACACCCCCAGCTTGTGTTGATAGTGG  
GGGTTGCCACAGCCCAGGTGCAAGAACCTCATGAGGTTTACCTGGACTCACTGCTGGAGCC  
TGCCTGGTCTCACTGGATGGCACCTCATCCCTCTGGAGTGTCACTGAAACCCACTGA  
AGGAGCACCAGCCCAGTGTCACTGTGATTGATGAAGATGTTAACAGAGCACCAGTGCCA  
GTGCTGACCTGACACCACCCATACCGATTGCAATTGAGCACAACGCTGACCCGGTGCCACTGAAATTGG  
AAGATGACTTCAGGTGGGTGCAAGATAC

#19

TTACATTATATGTCTATTTCACTGAGAGAtgtCTGAAGACAGAAATTAAAGAGAACATCCAAG  
TTATTGTTAATAAGTATCACGATGTCACTCCCAATTCTACATCTCAGGA  
GAAATTAAATTGTCATTAATTCAATTCTTAATGAAATAGCACTTGGAGAGCAGCTCATAT  
GTCTGCATCTCACATTAAGCAAATAATTCAATTGCACTGACGGCA

#20

TCAAGTTAGCTATTGAATTAGTCCACTAACCTAATTGGAAGGAGATGGTATGTAATTACCTTC  
CACATCAAAGCCAAGAACAT

#22

AAtTGGACGCTGGAAAGAGCATACtgggagAAgcaatgTGCTCACTTACTTGTCTCATGCTTT  
TCCCTAGGCACCCCTCTGGCTACCATAGAAAGcagGATGCTGAGCTAGATGAGCCTTAA  
TCTGACCAAATGGGtaTctGtaAGTAAATTAAACAGTGTAGCTTGGGACCTGGAACCTG  
GTCATCTGTATCAATAAAACTGATAGACCCCAGCTGTGAACGTCTTGTGCAAGACATTCAA  
TCAGGCTAATTAACACTCCCTGTACAGTTCCTAGGCACAGTTGAAGAGTTAGGCTATGG  
ACTATTCCCAACAAGCAGTACTGATGAAGCCTTTAATGAGAAAGAGTTaaaacaGTaaGa  
aGAGAAACAGCAAGCAAATGCCTCAAAACTACAGGACAAGCTGGTAACATTGAGCT

TAGTATGATGAAACTCAGTAGGACTGCTTTATTACAGGTAGAAACTCACAGAATTAAAA  
ATATGATTTCTTCTCCCCTAGAATATCTG

#24

gactGtCTTACAGCATATAAGCCTGGTACTGATGGAGgCaagtaAtGCTTGCTCTCCTTCTCA  
CTATTGTCAATAATGTATGATCTCCTCCATAGCAACAAAGGCAGTGAaaCACATTTAATT  
AGCACAGTTAAATGTAGCTTAATCATGGTAGCTAACATAGTTAAATCAGACTTGCT  
ACTGAGGGCAAATCATTAAAACCTTTCAAATGCCATTGATTTCCATTGTTGTTATGGC  
TTTGACCTGTCTGGTCACCATCCCTTGGCAGTCATGGTACCATAGCCATCTGACAGCT  
CTGTTCTGTCACCAGTGGTCTCATCACTGTAGTACATGAGAGAGGTTATTGCTGAAACCA  
AGTGGGACATCACTAAACGTAGGAAAGCTGTTATTAAATGTAAAATTAAATACATGACT  
GTTGTGTTGCTGTGAGTGGATTTAAATGTTAAGTGTGTTAAGTGTGTTGCTTTTAGT  
GCTTGAATCAGGTACTAGATGAGTGTGTTGGAAAATTAAATTCTTGTCTATGGCCAG  
TATTACTGCATCAGTGCATCTTATTCCATTATTCCAACCATGTTGATATTCTCTTAG  
GGTTGTAATTAAGAATAGAAGAAAATGAAGGAAAATGAGAAGAACGAAGAGTTAG  
AAGAGAAAAAGAAACAGGAGTACTGAAAAAAGAGAAGGTCAAGATATAGTAATGCC  
CTCAAAATGGCATACTGCCGTAAATAGTCATGAGTTG

#26

gtccATCGCCTGTGCAGGGGGCCGGGGTGCAGAAGAagaTTTATTCCCTCAAACAGAG  
GGGGAAAGAAACCAGCACACGATCCAGCTTGTAGCTTAAACCCAGGCTGCgGGGAGC  
CCCGCGCTCCACTCCTCACA

#27

gAgctGCTGCCCGCGGGGCTGTGTCCTTCCCTGCCACgCTCACAGGAACCAACTCTGCGC  
GGCACAGGCATGGACTTGGTCTGAAAACCTAAATACACTCCATTAAATAAGAAGTA  
CAATAAATTGCCACGTTAATATGCCCTGGGTGCATCATTTAAATTAGGCGGCCTC  
AAGATAAATAGTGTGAGTTCAGATG

#28

GgaGAGGATAACAAAACCACCAACCCAAATCTTAATAAGAGAAGTGTACTTTCTCAGAG  
CACTGGCCAGAAATCAAGCTCCTATC

#29

CTGCAGGCAGGGAGGAGAATGGATTAGAGTTAGCAATAATCTGGACAAATCAGAGGAG  
TGCCCTGAACAGAATGTGAGGTAAATGAGGACAAATGCAAACCTCTGTGAGGTA  
GGAACACTAACACTATTAAATGAAAGAACAGGAAATTACTGGCTAACGATCCTCCTGGAG  
AGAAGTATCTGGGTCTCT

#31

ATCCCAAGGCTGACAGCTCTCGGGCGTTGCTCCgTTTCTGCCTGCTGCAGCTCCCTG  
CCAGCACTACCAGGGAGTTCCAAGTGTCAACCACCGTCTTCCAGAAGAATTCTCACC  
TTACAACACGAGGCACGGAGAGTTCTATTAAAGTAACAGCAGCTCCGGTTAAGGTTA  
CGTACCGAGGCAG

#32

CTGCAGTCTGCCAGGAAAGATGAGGTACCACTAGTGACATgTCAGGTAAAATTC  
CAAATCAGATAACCCAAATCCTGTTGGTGGACAAAGCCATGCCATAAATAATCTGA  
ATATTGAATAATGATAGTACAGTGATCCTGCATAGGAGATGGTGAGTGGCGTGGCGT  
ATGCTGCCAAGTGGTC

#36

TTTTGCAAGCACTATCATGGCTGTAGAAAAAATAAAACATCTCCATAAATACAAGCCAG  
CTATGTGGCTTCTAACACAGTAGCCAACAGCCTCTGGTCTGAGTCAGGGAGAGTAGG  
CTTGGGAGAGTGTACCTCTCCTACCCCTGCCCTGGTGAGATAACAGCAGGAGTGCACAA  
TTCACCTCTGAGCTTCAAGTGGTAATTCAAGGATTAGAGGAACGTGGATTATT

#37

cCAGGATATGGTTGGCTCTGGGCTGTGAAGGCNCATTACTGGGTTGTACGACCCA  
GGTGCAGACCTTGCCTGGCTTCAACCTCATGAGGTTCACCTGGGACCACTGTTG

AGCCTGTCTAGGTCTCTGGATGGCACCCCATCCCTGGGTGTTGATTGTGCCACATT  
TGGTTGATTGCTCTGCATTCTGCTGGTGCTTCAGCCCATCACAGATGCTGCCCTCATT  
GAAGATGCTTCCACAACCTCAATGCAGGGTGCCTTAUTGCACGAGGCAGGTAGGTA  
ATGCTAAATGCTGTGCTTATGCTCTGTGATGCCCTCTGGTCTGTGGCAGCCTCTGT  
CACCCAGGCAAGCTGTGATGACAACCGAGGCCTGAGGAGCTCTGGAGGCTTCATTCTT  
GTTTCCAGCCATCATCCAGCTGAGAAGCGTGCCAAGTCAAACAAAGCTGTGCCAGAAGG  
TTGTACAAATAACCAGATCTTATTAGCCTGTATCATTGGCTGGC

#38

CAACATATCGATCTGTGCAGCCTGCTCGTACAAgAgTTCATGAAATA

#39

tGAATTTGTTAACCTTTGGTGAGATTCTGACTTGTGAAGGCACAAAATATAGTA  
AATTAGAGTCAAATAGCGGAAAAGTAGATTGTGAATGTAAGTGTGAAAAAAGTTGTATT  
CATGAAATCATATTAGGGTAACATCCACTATGTATGCTTATTATATTGCAATTATAAC  
AAATGAACCTCATAAAAGTACTGAATCTCTAGTAAGTTAGTGTCTTCTACAGTTGCCTT  
TCTCCTATTAGTCATCCAGACAGAAGCTAAAAAATTATCAGACTATTAAATTGTAGAT  
AATGGATAGCAGAAATTAGTATTGGAGGATGAATCACATAAAATGCTTATGAATT  
GTTCATATAGTACAGCCAAGAA

#40

tAaGAAAATGTTGAAAACTTGATGTTGgAGagaatTAgaTATCTTCTTGGCAGACAGC  
ATCCAATAAAATTCAATTAAATCAGTAGTATTAGTTAAATGtCtTGGCCAACCTCTGGAG  
AAAAAGGTTCTAGTGAAACACATACAAATCATTTGGAACATTCTCTAAACCTGTG  
ATAAAGCAAATCTCAATTTCAGCATGACAGCTGGAAATTAGTTAGTCTTCAGGCTTA  
GTTCTTAACCTTCCTGTCACTATTAAAGTTCC

#41

GCTAGAGTATAAGTGAGCTCCATATGTAGGCAGGGAAAGTGACACTGCAACCATAAATAA  
AAATTGTTCCAAAACCTCTAAGATTATAGCCATGCTCATTACTTGTCCCTCCTG  
GCACATGTGAATAGATTAAATTATGTTTATTGATTGATTGATCTTATTGACGACCAA  
GCTGAAAGCCAGTTCTCTTCTTGAATTCCAGTACATGCCTGCTAAGTCATTAA  
AAACCCCTAGAACCTTGACTCTGCTCCTGTGCTTCACACTCCCAC

#42

TGTCTTAGACACAGTAGCAAACCCATTTCATGACACGTCTGTGCTCCTGGTAgACAT  
GAACATTGTTCCAAAGTCAAACAGATAATTACAATGCTTAGGTATTGACCTGTGA  
GGATGAAAGACATTGCACTGCTTACAGCTGCTGCATGGCATAACCACGTCCAATTGAAAG  
ATCTGAAACTAACTAAGGCAACTCCTAGAGTTCCAGGATCATTGTGTTTCTGAA  
ACTATCCGTATGTAATTCACTGATGACAATTCTAGCTCATGCTCCAACCGAGTTTCT  
CAGAAACATTGAAAGAAAAATTAAACAGAGCGGGAAAGGGCAACCTCTTCATCCT  
cAGAC

#43

AAAATGAACACTTAACTTAAATTTATGTTCATTATTCTCTATATTCTAGCAGCT  
AGTGTGTTGTGAATACACAGCTCTGCAAGGAAGAAG

#45

tCCACAAAGCCTCAAACATAGCGATCTGCACGCACAgCTcTACAGCCTCAAACACAGTTAC  
CTAAACACCGCGCCCTCATAGCTGT

#47

TTCATGGTGAGAGATTGATGTCACCTTCTACACTTATGGGAATGTTGTTCTCTAGTT  
TAAAATTGCGGCCTTGTGTTGATTGACCCCCAGCTGATGTAAGCaCATTGATAGAC  
ATATCAGATCTCAGTAATTCAACCACTGAATTGGCTGTCCTCCCCACTAACCTTAAGAA  
AATGAGTATGAAATTATAATGTGGAGAGAAAGATAATTATGACAGTGCACATTTGTAG  
TTCCTCTCCTCCCTtGatGAAGTATGAAAGTTCATGGAACGTGCAAATAATCTCAGTTCA  
GCATAAAATATTCTCAGGGCAATGCCATTGTTAGCAGTAGCAGAAACTGGCCAATGGGAG

TGACACAgANATTTAGTTAAATTANAGTCTTAAATTGTTAGAAAATTGCAATAGC  
TTTGTGTCGTGCTGACATCTCATCTAAGGACAT

#48

CAATTATTTAGGagaGGGGTCAGTCTTCTTCTGTATTGTTCCAAAGAGTGGACATGCA  
TGCTTAGTGTAAACTACTGTATTCCAATGTCAATGCAAAACACATTGGTTGAATTTTTT  
TTTTTTCTgATTCCAAATTGCAgCTGTTTGTtGgACTTTTTTCTtCtAtATCCTTTATTT  
TTTTtCCCCtGATTAcAAAAGGTTTgANAAtCGGGtAGGTtCaCCaCatCAAatcgAGGGGctgTTtGt  
ctGGgtCGtacAgTTtCAGtGcatGGctGGcAAGtAAGGcaCNAAGc

#49

CCAGGGGTATTCGTTGCAAATGTCAAAGTATTCTggTTGAAGCTGAAAAGGGTTGCAG  
GGTGCAGAGAATATGAAGGTGGGAGGTATTGCCGTGCTAGAAAGGcatTTtaCTCCTCT  
CTGTGCTGTTAACCACTATTACCTTCAAAAAATTGGCAGAGTAATAACCATAact  
GTCGTGGTCATTGAGGAGCATCAGTTCTGTACAgAATTAGCACAGTGCCTTACTGTGCT  
AAACCATGTtGtGtGcTAAATTCAATTGCTGCatGTGTGTACGGAATGGCCGCTGACTCGG  
TTCTTGTCATGcAgGcTgAaAACCTGGCTGtCTGtGgATGAAATCCAATTCCACTCggAG  
cTCCcAgGgaggcCCAgcCTTAgTGTAGtAACTCACTTaCCACTAATGcGcCCAAAACAATTTta  
GAGcTTAAAAtaCagAcTTCATACATAATACATGGtACAGCaCGTGAATATtcTGTGGGTAT  
TAAAGCATGGATATTtGCACTGAAcAGaCTCCTGtCtAAAAAGcATTcACAAATTAAAGAC  
CCGATCCAGCATTGAAATTGgAAACTTCAATggCTTCTGGGAATTGTGTCAGGAA  
TGCAGTATCTAgCCTCACAGTTGGATAAAACAGATGCAGCACATTCAAGTAaCAAGTTGAT  
AGCATGTACAGTCatAAAAAatAtTGTATAAGAATGgAACGTTGTAATCATTtATACCCt  
TTGTATCtAAGATAcTctGACTCATGGtGCTTCACTTACTGACTGAAATAAAATTGGCTTc  
TGTTaCCTTGGCGGTGCTGACACAATTGGTACAGCAAGCTGAGGAGTCTTGGATCT  
TCCATTTCCTTACTAAGTCCAGCTGAAGAATTAAATTAGTATTGAGCCAGAAAGAA  
AATAGTGTGGCTAAAAATCTCAGGTCAGTTGAAGGGTCATCAGTCTAAGcAGCCTCT  
TTTATTGAAACTGAGCAAATGGCTACAGAAAATTAGAAGACAAAATAAAAGC  
GTTGATGttCAAGTGAAGGCCATTAGTCAATTCC

#50

TTTGTCTCTGATTGATTTCAAGGCAGTGAATCTatTGTAAGTATTGGTCAGTAGTGACTT  
ACTATTTGCTAATTCTCTCCTCTGTGTCCTCCGCTGCCATTGAGCCGCTGTATGAAAAG  
TAACATCAGGGTGTCCACAACCTCTAGCACAAACTTCTATCTGGAAGTGATACTAATAT  
TACTTGTGCTTACTCTGTTCTGCTTAATTAAAGTTACGCTGCCATTGCTAGCTTGTGATCT  
GGATGTCGTGAAATACTATTCTAAAACAAATTAGTAGTTAATTAGTCTAATTAGTG  
GTAGACGTGTGCTTTCTTTGTATTAAAGGAGAAAGTCTGGG

#55

GATGAGCTGTCTGGCCTCAGATTGGACAGGCAAGTAGACTTCTACTAAAATATGAGT  
CACATTAAAGACTCGCTGTACAATAGTAAAGAAAAGAATATGAAATTGAGAGGGGTCTT  
CCTGAGTATTAACTGGATTCTAACGTCTGCACTGCAAATAATGTGTATGGGAAC  
TGCTGAGTCACAATCTGAACCACTGATTGAGCACCTGGGGAAAGGACCCAGTCACCCC  
TGGGAGTACAGGTGAAGGTAAACCTATGTGACCAGAAGGGGGTGGAGCCTGGCTGC  
ACCTTTCTTAGACCTCATTACAGGGCTGACTGCCACTGAGAAAGGATCTCTTCTGGAGG  
TTCTTCTTAGTGAAGCTCTCTGTGAAACTAGAGTCTCCGATACAGGTGAGCAATT  
TCTTCCTCATTATAGTATTTCTATTATGTTAGGCCCTCATCATCACATGTCTGTTG  
TAATGCTTTATCATACTGATCTGTCTAATTACTACATAATTATATATATTAGACAA  
GGAGGGCAAGGAGCTTATAGGTGGAACTAGAGCTAGTGGAAAGTCTCAGGAAAATAGA  
GAGTTACTTACCGGTCTGCTTTCAGGCATCTACAGGAACCTTGAGCTTGTGTTG  
TGAAGATAACCCCTAAAGAAATAAAACTAGAGCAACACTCAGTGAAGCAGATACT  
GTACAGAGATATTCTAAAGTTAATGAAATAATCTTTATTCTCAGTTATTAGTTGCGTC  
TATAATTAACTTAATTCCAATTAGTGTGTAACATACTTTGTGTATTGTAAGG  
GTGAAGGGTACTTGACTTAAAGTAGGAAGTGAACAAACAGTTACGTCTTATT  
TGTATTAACTTAGCTTACAGTGTGGCTCATGGGGAAAGGAGGGCACTACTGAAGTT  
AATATGCTTCTAAAGCTTATTGCTCCTTCAAATAGTGTAGTTGGCGTTTTAGTCAG  
TGATCTGCTCTTGAGCAGAAAATGAAGCTTAAACAAGAAGTCAACTACAGAGGCTTA  
GAGCATCTAAAGAATGATGGAGGAGAACCTTAATCACTCTTAATGTTAGCAGAAATCTTATA  
TTTGTACAAAAAAGCAGGTGTTGGCTGCTTGTGATGTCATTACCAAAAATCGCTT

AATTIATTGTCTGAGGTAGTCAGAGCAATGACTCATATTCTATACTCCTTACATTGTACC  
CAGTTATGTAATTCTACACAAATAATGCAGAC

#56

AAAGCCACATGGCCTCCTCTGCAAACACCCCTGACGCCAAGATTCTAATTCAAGTTGCC  
AGCGTTGGCAGATGACAAAAACCGAGCTCCTTAGAAAAAGCAGAGAAGGCAAAGTT  
GTTCTCGAGAATTCCGAGAACTGTAAAGCACAGCGCCCCAAAATGAATTTTAAA  
TGATTCTGCCAATAAAGTGCTACATTAGCGAACCCCTTAAAGTCTGCCGTAGGACGT  
GAACGCAGCTGCGTATAAGTGGCAGACTGCTAAGCTAAG

#57

CTAAATCTCAGGTTACAAAATGAAAATGGGAGTGCTACAGTAAAACCCCTGCACACGTG  
GAAGTGATTACATTGCAAGCTTCCTTGTATCCAGTGCCTAGAAGGACAAGAAATCCTG  
CTGCTGGTCTCTGAAAACCCAAAGTCCCCAAATTCTTACTGGAGTCACAGCGCTCCTT  
TGAGCACCATGTGCCATCATTCTCGGGCCCCGCTGGATTAGTTAATGAAAGT  
CAAACCTTAATAAAATACCTTAAAAGGAAAGAAACTACTAAAGATGGTGTTCATTTC  
ACTGCAAGCAGTCTCCTCTATTAAATTGAAAGTATTTCACCTGATACTTGGAAA  
TGAAGTCACGACCTTAATAATTCCATTCAAGACTAAATGCATAAATAGAGTCTGTCCAAG  
TGCTAACCAAGAGCATATCCCATAATCTGAGTCATTAAATGGCTGATGAATTCAAGCTAA  
TAATGTCAGAGAACATTGTCAGGTTGATGGACCCAAACTGCCCCTTGTTTGACAGTAC  
TTCTATGGGAGAGGGCTTATGGGAAAGCAATTACCCAAAAGAGATTGGTTAATAGC  
AGACTTCTCCAAGGCAATTGGATGGATGGCAGATGTAGATGTATATAATGGGATTCTT  
TTTTATTTCTTATAAGAGCACAAATTATGATTCAACCTTAATTCCAGTTCTGGTATT  
GCCAGTATTCAAGTGTACATCCAATATCAGTACAGGGCTTATGACTCCCACGGGCATGA  
GCAATAGAACAGAGAACGTTCATCCAAGGTTGTGCTGTCAGTGCACATCCAAG  
TGGTGA

#58

ATTTTATACAAAAAATTCTTACTAAGGGAGTCCTCTTTGTAGAAAGAGGACATTCA  
CAGCATTACCTGGTGTGTAAAGAAAAATAGGAAGGGAAC

#59

TTGCGCAGTTGTACAGCCTCTAACAGCTACCCCTAACGCTGACTGCTGTACAGGTCAAG  
AGTGGAGGCTATGTTACACATTAAACAAGATGCTTGGGTACCAACG

#61

TTGCTTAGCAGAAATGGCAGTGCCATCTGATAGTCATGATGTATGCCGTACTGAAAAT  
TCAGCAGCTGGATTCAAGATAGATAATGAGACTGAAAGAACAGTGTAGCTAAGAAATCAA  
TGGGGAAAACAGCAAAATTCTTATTGCTGCTCTTGTCTTATAGTCCCTGCTTCA  
TTCTTCTTAAATCCTGAGAAGCTCTACTGTAAAGCCTTAATGGTACATATGGCTCTTAC  
TTTATTGGCTGCTGAATATTGCACTCTCTGTTACCCAGTGGTTGGGAGTGCTTTTA  
CATCATCAGGTGTGAGAAGAAGCTAACAGCTACAAGCAGTTCAATTCTGAACTGTG  
ATATGCAAGATAATCAAGATTACTTCCAACAGTAAATACTCTCTGATTGTCAGGT  
ATGGTTATGAAGTGACAGGATCTACAGTAAATACAAAAAGAAAAAAAAGAAAAG  
AAACAGAAAACCAC  
CAAAAAAAACCTCTATTATTACAATCAGCTCATAGTCCGTGATTAGATGGCTGACAC  
TTCAGTTAGGTGTAAACACAGATGGAAGAGTACTCGTGTGCACAGTAAGTCCGGATTCA  
TGAGTTGGCAGAT

#62

AAAGAGCACAGTACAGATAGTGCTGAAGCGCAGTTGAAAGTACACTGGCAATGAGTGA  
TGGTGTGATACGCAAAAGGTCTATATTACATGTTAAAACCGGAATTGCAACTTTGAA  
AAATTGATTAGCTGCCAAGTCTGTTCAATTATCTTGTGCTTAAATTACATGGAAAG  
AAAAAGCAGGAAAGACGGCAATAGCCACTTATGACTGGCATTGATACTGCTCGTGCCT  
TTGAATCTGACAGTTCAAGTTCTGAGCTTGTAAATTAGCATTATTATGACATAGAAT  
CTCACTAAGAAAAGAATAAAGTAACCACTATAAGTTGACTGGAGTAAGTGTGTTGGGA  
AGTTCAAGTCTAAATGGCTTCTCTCAAATACGTCTCCAGTTGTTAAAAGGGATG  
AAGGCAGCAGACATACTCTGTGTTATCAGTCCTCT

#63

CTGTTGTTGCTCTACTCTGCATTCTGAAGTGTGGCTTGACTTGCTGAGAGTAT  
TGGTAGTATTCAGAGCAGTGTCTGCTGCTGAATACTGAGCAGGGTATTGGAGCTG  
GGTGCC

#65

GAGATGAATCAGCAGGGAAAATGGCCGGGTTGCCTTGGTGGAAATGGTAAGGGCA  
GGTACCGGGCATCTCCTTTGCAGAGCCCTGGTAGTGGACATCAGCAGGACTATCTCA  
GCTGGCATGGGAGGTTCAAAAGGAGTGCATCCATGCATTAGCTGAGTTGGATCTT  
TCTCTGTAATCTCAAAAATTACACTGTGGTCTTCATTGCTCTTCTATTCCCACGTTC  
GCTCTCTCACTTACACTGTGGTCTTCATTCTCTATTCCAATGCCCTGCTCTCAATT  
CCTGTCTCCAGCACTGACACCTCTTGTCTCATCTCCATCTTCTGTCTTGTCTATT  
GACTGCAAATTGTTGATAAAGGGACCATGACTTAATAGGTGTTGCACTACATCTGGCA  
CCATTGGCTTAATCTCAGCTGGACATGAAGTCTACTATTAGCGAATAATTAAATA  
ATGATTAGGTCTAAAGCAGCAATGCTGCGAGCATGTGATACTCCACTAACCTATCCAC  
TGCTCCTATCACTCTCTTCTTAATGCCTGCTAGATTAGCAGGGATTACTTCAGA  
GCCACGGCTACTGAGGGCAGCCAGCACGGCTGTTGGTAGCCCACGGATGCTTCCCC  
CTTCTGAAAAAAAGGAAGGTTAAAGTGGAAATTAAAGCTAAGTGCAGTGCAGGGTCTT  
ATGAAAACCTAGAATGCATGTGCAGAACGATGTTATGCGTAATATAAACAGCTG  
TATGTAAGGGAGATCACTGCTCTGTTAGGAAGTAATAGCTAAAGAAGCATTTACTTAT  
TTCTGACAATTAAATGGACATAAAAGCTTATGTATTCTAAAGAGGAGAAGATTAAAGG  
AGG

#66

GTTATGTGACTGCCTTGGAAACAAGAAGACACAGGAACCTACTTGGCTGGACACCTCTG  
GAGATGTCCTTGCTCAGAGGTTTGGGTCTTCACCCTGCTCAGGGCAGAACAAAT  
CCCTCAGTATCATTCTAGTATTGTGCTTTAAATGAATTATCTTTCTTCTTCTTC  
CCAAGGGCTATGAATATGCTGAAGAGAGTAATGGAAGCTGTTCAAATAGAACAAAACA  
ATGAGGCAATTAAATGACTTTATACACTTGTACAGATAATTGACGTTGATTGTCAT  
TTAGGCAACTGAACAAACACACGTAAGAAGAGATGCAAGAGAAATGTGGAGTCACCAGCCCTGG  
TCGTGTTCTAGAACCATGGAGATGTGGCACTGAGGGCATGGTAGTGGCATGGTGG  
CTGGGTCAACAGTGGACTTGGGATCTAGAGGGCTTCCAGCCTAACGATTCTATG  
ATTCTATAATTCTATGATATCCATAGGACCCGTTTCCCTAACGAGGTTAGGTGTTA  
GGCTACAGCACTAATTATTCAATTGAAATATTGCTTCAATGTTGCTTCAATATGG  
TGCTCATAGGAGCAACAGTCTGGAAATTGCTTACAGGCCCTAACAGTGTACTGG  
TGCCATGAGCCAGCAATGTGACCTCCTGAATCCTAACCAAC

#71

CTGCAGTTCTTAAACAGCAGGGATGCCtGTTTGGAAAGAACTTATGGCTtGTTGAAATA  
C  
ATCTTATGTATTGCTT

#74

GTAGAATTCAATACGCTGGGATGCCAGAACCTGTAGATGAGAAAGAAAAGGACTATGG  
AGGAAAGGACTGAATTATATCCAGGAAATAGAACGCTTATATCTGAAGCCTGACTATGT  
ACAATTCAAATTACAGATAAAAGCCTGAAGATTCCATAAAAAGAACGGCTAGGGC  
TTCCTTGTCTGGTGCCTGTTCTGGTAAGTAAATGAACTACAGCTGGTGAGA  
AGGTATGAGTGGTTCTACCAGGTTTCCAGAGCCACACTGTTGCCATCTGCTCCTTTT  
ATCTATCTGCTTTATTACAGAAAAT

#77

ACTTGGGTTAGCTGTCTGCTATCAGCAACATTACTATTGCTCTCTCACTTCGA  
GGATGTTATCTTCTACTACATATGAATAAATTCTCGTAGTCAGGTTAGATGGCTT  
CAGATCACCTGGAATTCTCTTGTGGTAACCAGAAAAACTGCAATGTTGCCAAACAA  
CAATTATTGCTTCAATTCTCATTCCTGAAATCTGGCAGCGGATTGCCAACGCTCCCTCCG  
TATCTCTCAAAGAATTAAAGAATAATTCTTATTATTATTTATAACAGTAGAAAG  
ATATATCATAGTCAGTCAGTCATTCTGGATGCAGATTGTTCCAAGAACAAATT  
AACCTGAGTTCTGGCTTGTATGCGCATCTATTCAATATCCTCTGTTCCATTGGACTCT  
GCTGGGAGAAATCTGTGCCAATGTCGATTGCATCATGGCAATTCAACTTGGAAATT  
GTTGAAACAACCTTATAACAGTACTAAACACATTCTCTGTCTGCTGAAAACACAAGAG  
ACACCTACAAGCAGGTGGATTCTGTACTGTAGTGTATTAGCTGGCATGTTCT

AGACTGCTGTTAATTCAAGCTGTTGCTCAATTCACTGACAGCATTAGCCAACATA  
AAACAATGAAGACCATAAAACATCAGGCCACTAAGAGGAATACAGAAGTCACTTCTG  
CAATCAGTCACTCTGTACGTGCATCTTGAGTGCACTTACTGTGAAAAGCAATAAA  
TCAAGTAGAGCCAGTAAATCAAGGCATATCAAGGTACAACAGTCAGCATGGTCACATGC  
TGGAGTTAGGTGGAGTTAAAATCACCAAGTCACAGAACCCCAGCAGTGTCAATGACAT  
ACAAAGGAGAGAATCATAATCATAGAATCACAAATGGCTGGGGTGGAAAGGGACCTC  
AAGGATCATCAAGCTCAA

#78

AGTTGAAGGAAGGCCCCAGGGGTTGGTGTGACAAGGATCGACACCACTGATGACAGGGC  
TGAGTCTCCTGCTGCCAGGACCTGCTGCCGGGGCACTCACATGCCCTCCCCATGCTCAC  
CCCCACTGCTTCTATATTGGCAAAGGCTGAGAGTCCCTGCTGAGAGTCTTGGCCTGCAG  
AGCAGTCCCTCCCTGCATACTACAGCAGCATAGCAAGCCAACCCGGGTGCC

#79

ATTCCTTCCAGTCCCACAACATTACATTGATCAGGTATGAATTCTTAATTACAATGGG  
CACTTCTCAAACATGTTCTAAAGCAATATTGCTTTCAAACACTGTGAAACAAGTTA  
AAAATATTATAAAATAACCAAGTACCTCTGTTGTTCAATTGCTTAAACTGTTGCTCCAGC  
TACATTAAATAGGCAATTGTTAAAACCTCCATTCCAGAGCTCCATCTGCAAGCTAA  
AGTCAGGTCTCAGTTAGAGAGCATTGCATGCCAAACCTCACATTACAAGAGTAG  
GAAAAGCTGCTCTACAAAAACTCTCACAGTCAGCAGCTGCTTGCCTACTAATACATT  
TCTTACATGAAAAAGCCTCCAGCAAAAGACTGCACTGACAGGATTAACCTGCCCTCC  
AAAACACTTCACTGTTCCCTGCCACCAGAGTTCTAAAT

#80

ACTAAATGAGACGGACTGACAAAGGACGTCTCCAATATCCACTATTGCTGCTCACATGA  
GAAGTGTGTTGGGAGGAAAAGGCTTGCCTGGTTATCGGTTACACTGTGTAGGA

#81

TGAGTAAACATGTTCTGTTTAATTATCTCCAAAATGAGTGTCAAAACAATATGA  
AATCTCTATATACTAACATATGACGTAAGGCTCCAACCTGGGCCCTTACTCTGTAAG  
AGCATATGTT

#82

TGCACTGAAAAGATTAAGAAATCAAACAAATAGGATTCTTAATAAAGGGATCTTAAAT  
GTGTTTAACTCCCCCGCAGCTTAGAGCACAAGCATTCAAGCACTATATAGATTAAAAC  
GTCACTCACTGATAAAACTTAAGAGCTCGTAGTGTATCTTTGCTAGCTAGCGATAG  
CGCTCCAGGTTATTGAAGGCTCTCGCATTACTTCGATGTATCTTACTTCCAAACTAA  
TGCAGAGATTCTCCTGAAAATAAAAAAAACACTGCACCTTAATACTAACAGCTGAAAAAA  
CTGCAAAACTCTCTCATTAATTATCTGCTTAATCTAACAGTGGATAAACGTCCCTACAT  
TAATATTCTTACACCT

#83

AATTACCTGTGCTGAACAAGTTAATGATACACATCTCATCTGAGTTAAATTCAAATCT  
GAATTAAACTAGAGCACTACAACACTGCACTTAAGCTGAAAACAACCATTCAACAG  
GAAATTCCAAGGTTAGACAAAAACTTCTAAACATTCAACAGCTTCAAGAAAAAGA  
CACAGTGTGTTATGTTGACTTTAAATCTCCATCTAACACTAGCACTCATAGTTGCATT  
TTCAACCAAAATCTTTAAAGGAGTGGAAAGAAGAACGCTGTATGTCTCAGCACAAA  
AAAGAGCTTATTAAAATTAGATTACAGAAAGATCTAACAGCAGAACAG

#84

TGAGTGGGTGCTGTGAGATTCTCCATGTGCCTGGAAATGCTGTAATAGCCTCACACAT  
GTGACAGACTAACGCTAGTATGCACTTCACTGGCTTGTCTGCTTGTAGCTTCTGGTT  
GGTACCTTGCAAGCAGGAATTGTTCTGCAGGAAACACGAA

#85

AGGCTGACCTAGAACAGAGCTACTCTGTGTCACTCACAAGTCCAAAGAACAGGTACTTAG  
CCAGAGTCTAGCTGCTTTTGTCCACAGCTGCCCTGCTCTGGGAGAGGAGATGCAGCTC  
CCCCAGGCTCTCTGAGCAGACAAACAGGGCTGCCAGGGAGCTACTCATGGCTTCA

GGTACTGGACAGAAAATCCAACCACATAAGTGTACTTATGGTGAATTGGGTACGTTTT  
GTAAGAAAGCTGCCCTACAGCACTGAGCATGTTACGTGAGCTTTAAATACAGGTGC  
CTTCAACTCTGCCCTTTCTCATGCACATAAATACACACTACGGTGTCAAGCCGCTCA  
ACTGCTGAATTATGTTATCATTCTAATTGATCTTTGGTGTATTCTGAACGCCCTCC  
TTTACAAAGAGTCAGACTCACCTTAAGAAGTATCAAATAGCTCAAGGAACCAAAT  
AACAAAGACTGATGTAACATCCCCCTTCCCTACACTGAGCTGGCACTAAAACAACA  
GCAGTTGGCACCTGGAGGTGCACTGCCAATACAAAAGCATTACAGTT

#86

TTGTAGAAAATCTGTCATCTTACATTCTAATGCTATCTGATAAAGGACAGTGTAG  
TCTTACTAGTTGATAGGTTACTGTTCTGTTGAATCTGATTACATGAGAAATAATGTGC  
CGTGAGAGGGACTGAACGCATGACTTCATCCTCTTGACACCTACTCTGTTGATAT  
TGTGAGCTTATTATTAAACTCCTGTCAGCTCAAAGAGAAACAGCTACAATATT  
CTGAGTATTCTAATGAATGACTGCTCCTTCCTGAGTCAGAGGGTTGTTCTGAGTATCAT  
TTTGAATCATTAACTTGTGCTTAATACAAGCAGTGTGTTGTCAGAGCTGATTG  
TGCAAGTTGCACAGATAAATAGATGCATACTTAAAGAAATATCTAATGAAAGTAGCT  
TTAGCACTCACTTGTGAGAGTCAAACAATGCCAAATCTGAGATATGCAGTAGAAAAG  
ACTACATGGCAGAAAGATATTCTCACGGGATTAGTTGAACAAGCAACACTTCATA  
TTCTCTGGCTGTATACTTGGAAATGTTACTTTCTCTTCTGTAATTACAGGTTAAGAA  
ATGAGCCTTACCCAGTCTTAATTGGAGTTGATTGAGGTCAATAATTAAAGCCAAAGG  
GAAATGATTTCTTGAATACCAGGATTGCTTAAgTCTCATAgTCAgTTATTCAAGtATG  
AtAaCCTACTGTNAAcATTTTaCCTCGcTATTgAAAaATAAtGCaATGCataTcagTTT

#88

TAACACAgACAGCTGCACTGGACGTACACCACAGCCTCTCGCTCAGTAGGGCTCATAGG  
TTTACTGACAAAGTGTAGGAAGGCAAAGCTTGTCAAGCAGCC

#91

CTCAGGGAAAAGGTACGGAAAGTGCAGGTATCCTGTGCAGGAGGCAACAAATGCCAGG  
TTGCAGACCCCTACCAGCAGGGCAGGAACCACCAAGTTCTGCTCTCACCCACGTCT  
TGAGCAGCACCCTGAGACTAACCTGGTCACGTTCCCTGACCTTCCCTGGTACTCCTG  
GCTACAGTGGCAGTAAGTCAACTGAGGGGAGAAGGGAGAGTTGCTGTTCCCACATGGGTCTG  
ATGGACTCGTATTACTCTAATGAAGAATTCTTCCCACCAACCCAGCAACGTTAGTGAGCG  
TTCACCTTGCCAACGGATAAAATACATCCCTTAAACTATCACTACGGGAGAGTGCTG  
CATTGGGGCATCTTGGAGCATTTGTTGATCCTTC

#92

GCAAAAGCAGAATACACTGAAGGAAAATTGTCAGTTCTCATATTGAAGTACAGAACGC  
TCTCAAATAGAACGCAGCCAGCCTCTCAACAATTTCAGCAGATCAGGCACCTCAGCAC  
ACAGTGGATTGAGGCACTGCTTGAACCTATTCACTGTCAGGAGCTCTCCACTGGTACC  
ACCGCTCTACATGAACACACAAGAGGAGTTCAAGGACATTCACTGCCACAGCTCCACTATTCC  
CAAGGCAGGAAAACACAAATCTAGAGCACCCTATGCCAGCAAAGCAAGAACAGAAAAT  
CAAAACAAAACCAGCTATGTGAAATATGAGACAACCTCCCTATGCCCTTCCATGCTTCT  
CTAGTCTTGTGCCATTAAACAGGGACAAGTAGCACCCTTCCACCAACACAGATG  
ACAGTACTGTACTCTTTAGAATCTTCAACATTCTGATTCAGTCAGAAGGCTG  
TTCTATCAGCCTCAGAGGCCCTCATGACTTTCTGATATTCAAGTCAGCAGATAACCA  
GCAAGTGAATGCAAGCTAGGTCTACAGCACAGATTAAAGTAGATGAAAGACCTTCAGG  
GAAACCAAGTGAAGTGTAGGAGAAGTCTTAAAGAGTCTAAATATTACCATG  
ACTCAGCAAATTCAATAGGCAGCACGAAAGCCAAGCTCAGTTATAGCCGTAGATCTT  
CCAAACAAAATAGTATGAAATACACCCCTTACACCATTCAAGAAATTGAAAAGGAGG  
AGACTAAAAAGGCCTACaaGCAATTACTCCCAGCTACaaGGCCTACcAAACAAACAG  
gATTAGCcTGTCTAaAAGaagcagtgttaCTGATCAAGTGCCTAAAGTCTTATGAGGTGCAC  
CACAGTAAAATGTCTATACAAACAAACGCACGCCAGTTGAAATTAGAAGACC  
AGTGTATATGTAAGTGGGGAAAACAAAATAAAACAAAATCAACACAAACACTTCTT  
CCAGACTATGTTATTCACTTAAACATTATGAATGTGCAGCCTCAAAAAGAAAACAC  
AAAGAAAATAACAACAAGAACACCACGCAATTCTAAACAGAAGGCAGTCCTTA  
CTGGCAGCTTACCCCTGCCTCTGGAGAAAGATCAGCATCTTACTTAACCTCCTCTT  
CAAAGAATATCTAAGAATACATCTGTATTCAAATCTCCTCTGATTTAACCTTCAGG  
GCTGAAGTAATAACTAAATTCTATGTTGAAATAGGCAAATAATACAACACTGCCAACGA  
TATCATCAGTACTCAGATGCTTATTCAATTCAAAATCGTGGTATGcTATTAA

GAAGATAAATAAGAAAATGAGACAGTCATCTTCCAGTCAGTTAAAGTGACTTTACTGACA  
ATCTGTAGGCTTGAATTAAAAAGCTGAACCATAAATAAAAAGCATCCTGCATCCTTCA  
GTTTCAGAAAATAATCATCCCATTAAAGCACCTTTTTGTTGTTAAATTACACATTAAT  
ATTTTATTCCATGTAATCTCTGCTTATAGCATTGTATACAAAtGGTCCAACTTGTGAAA  
ACTCTTAGGCCTGCTGTATTGCCCTCTAAAACAcAtGACTGCTGTTAGTCTGAAAAT  
GCAACTCCCCGTGCAGAGtctGGGCACCTTCACTCTGGGTC

#93

gaGTCAGGACCCAGGCTTGATGGCTCCTGGCAATcTGATTTATTGTCACCCCTGCCA  
CAGCAGAGGGTTGAAGTGTGAGCATGATTGAGGTCCTCGCACCCAGCCATTCTATG  
ACTCTGAGATTCCATGAAGGTGGTGGAGCATGATTGATTTCTGACCGGAGTGGTT  
GGAGCAGTTGTTCCCTCAGTGCCCACCTCCCCACATCTGGTGCCTGTCACCTCAGCAGTTA  
CAGCACTGAGACCTCCAGAGCTCAGCAGCTGGCAGCTTCAGCCTTGTGCAAACACTCAG  
GTACCACTGCGAGGCACTTCACTCAGATTGGAAGATCTCAGCCATGAGGGCTGTA  
AATGGACTTACATTAAAGAACAACTGAGGTTGTGAGGCATCAGGGCATGGCATCTTC  
AAACTGGCACAAAGCAGGCAGGTGAGGCCACATAAGCCGCAAACACATGACAAAAA  
CAAGTTATAGAAGTAATGAAATATTAAGCCAGAGAATTCTCAAATTAGCAATTG  
ATGTATGCATGGGTTGGCTCTGCTCTGTTATTCTGCCAAGCACTCAATTAGCACTTGAGT  
TAATAAGATGTTTAATGGCTGTTAAATTAAATCAACAACTCTGACACTGCT

#95

GAACATCATCAGATTGTTAGAAAAACAGCCTCACNGGAGAAAGCACAGTTATTCTCA  
ATGAACGGCTCTTAAC TGAGATTACGGTACGGACAGATTACAAGCTTACTTCTCTGT  
AGAAAGCAACAAAGAATGGAGCCCCGCTAAACAAAATATTCAAGCAATTAAAGATCCCA  
AAGAAAAAAAAGTCTCTCC

#96

GTCTCTCTGAATGTTGATGTTAATTAGCTTCTATATTCTTCAACTTGCTCCAGG  
ACTGGTGAACCTCACATCCTATGCAAACACTACCGAACCTGGAACGTGGTAGTCACG  
ACTGGCTCTATTACTATGCCTACAGGGACTCCTGGGTGAGCAACAAGCTCATATT  
CCTTG

#100

CATTGGGCTCATTCTCATTGTTAGCCAAAGACTTGGCACCCCTGCTGTTATCATCCA  
ATTTAGCATTACACATCCCTACGGTAACACAAACAGATCAATTAGGACAGTTAGATCA  
ATTTAGGACTTTGTGTTAACAGCTGTTAAAAGTTGGAGAAGGAAGGAAAAAGTT  
AAGTCACCTTGACTGATGTTAAATTCTCACTATAGGCAACTGCTGTTCCCTCTTT  
CTTGTCCATCACCCACCCGTAATTGAGGTGCTTACAAACACTACTCACATCCTGC  
TATGTAATTACATGTTAACAGCTGCATCTGAATAGGAGTCAGCTTGGCTGATCTTCTC  
TGTATGTACGTCAAGAGAATGCTAGCAAGGAATTCTCTACAGAAAGTGACAAGGACTAT  
TTACTTCAAAAGGGTAATTAAATTGTTACTACTGACTCGTAGTGTGAGAGAAA  
GGATTTAATTCTATGATGTCATCTCACACAGCAAATTCACTAAATAGTTATGATT  
ATTCCCATATGCTCATACTAAAAATGAAAGGATTGGAGACAAAGAGGTGCTGTATT  
GCAAAGCATTCAAAAGTGTGGGATTGTGAGTATGACTACTTCTGTTCT  
AGTGAATGTTGTAG

#101

AGAGtGAAGAGAAATAGGCTGTTGGAAAGCAGTTGCACTCCGTACCGAACCTCTGGG  
AATTCTGAGTCATGCCAAAGGCACTCACCTCCAGCGGAGGCATCTAAATTGGATGCCT  
GAAGAGACAGATTGTGAGTACCTCCAGCACAATGTATTCTGTTGGTATTGCGTTGTT  
TGAGTGAAGTCGTGGAAGTCTATCATTCTCAGATGTTGCAACAAAACACTGTT  
ATGCAGTTGGTCGCAAGAACACACTGACTGTGATGGAAGATGACTCTCCAACTGTT  
TGTTCCAAGCTGGGCATGGAAAGTCCACTCGCTTGCTCAGCCACAAAGGAGATAA  
TCCAGGTGCCGCAGCTCTGGATACGTGAGATCTCAGATCTCAGAAGGCAGATGAGAG  
AGCAGAAAGTGCAGACAGCATGTCACAAGTGCAGAGTTGAGTACCTGCAAATGTTATAAT  
TGAAACGGCTCTGGTGTATGGTAACCAGCAGGTGCTTACATGGCAGGCCTACAAAG  
AGTCTACATTGATCTTGATTGTTGAGGTGCAAAGACTGTTACACTGTCATTA  
ACTTATCATAGACTCCCATGTAGCAGTATCATAAATCAGCTCAGTGTTAGTAAATACC  
TAAAGCTGTGATATCCATCATCAGAGAGCTTACGCTGACAGATTGAAATGCTGTG

CCAGTCTATCTTCAGATATGAAATACATAATTAAACTCTGAATATCATAAAGCCAACAA  
GAATTAAAGCAgAAgCTGTATAACTTCAAACCTCTTAGTCTTGTAGTCTTC

#104

AGGCTCCTGCTGTCACCTGGTAGGCCCTGGGAGCACCTGCCCTGAAGAAGGACTTCTGAGGCTCGCTCCAGCCACAGGCACAGCCTGTACATGTGCGTGGCCCTGGGTCTGGCTGCCCCAGAGAGACCTGCTGTGGAGTGGGAGGTGGGAGTCAGGGTACAGACCAACTGTTGATCGCTTCAGCTCAGTTCACTGGAATGACTTTCTGAGGTCTTTCACTTGCAAGAGTGGAGAGCAGTGAGTCCTCGTCTTACGGCACTGGAGATCTCAGTGTAAAGAGGAGGCCTGTACACCACGACGAGCTGTGCTTACAGCTGCTACGCTCCCTCAGCTCCCTGCACTCTGATGAAGGACTGATGTTAACCAAAGCCTCTGGCTCCAAGAAGCTCAACAGCAAGAACTACAGGGCAAAGAAGGCAGAACTGCCCAAGAAAACCTCTGAAAAACAGAGATCACCACAAAGCTTGGGAGCCACAGCAAGAAACAGCCTGGTTAAAGGAGGGAGAGGGTGTGCGGGTGGCCTGTTCAAACAGCAGCATTAACATACCCATCAAATGGGCCCTTACAACCGAGGTATGTTAGCTGAGCATCAGCGTTTCTTATATATTGGAACACAATGAAGAGGtGTTCAAGGATTTTAATAGGGTAGTCTCACGGTAAGGAAAAGCTGTAAGGtGAGAAGGGTAGGAGGGGAtTATGTTACTGTAAAATACTCTACGCTGACATATAGTAAGGCCATCCCCAAATTCAAGTCCAAACTGCCCTCACTGcATTCTCGGGTGTGAtGcAgAAATTGGGGtGCCAGCATAACAAATTCAAGCTCCTACAGTTCAAGAAAATTGGGtGcAccaAGAGAGCATTCTtAaCATTGTCATGTAACAGCCTCTAGCTGAAGAAAAGATCTCCCCaAGGAGCACAACGTGTCACCTCATCCGAGTGTGTCAGATATCACaCTGCTTTTGACAAAGCTGGTGGCTTATTAAAGATAATACTTCCTACAAATTCCATTAGTCATAGCTTATTGGCTTGTCCAGCTTGCAGTGTGACTGGAAGAACGGAGGaGacagATGAATAGCTACATcTTcTCAGTGAATTGAAGAAAAGTGTGAGAACAGTGGATTGAGCTGAATAAGCTGTGAATAAgAgCAACTGcCcAGGCTTCAAGGAAATTGATTTTTTTAATTAAATTTCATAAATTAGGTAGTATACTCTTACAGTTAAAGATGCAAATCATAGAATCATAGAATCACCAAGGTAGAAGAGAACCTCAAGACCATCCAGTCCACCGTCCACCTATACCAACGTTCCCCACTAAACACATCCCAGTATGATATcTAAATGTTCTAAACACAAAGACCTTGCACAGGGGAATTCTGTGTTAAATCCCCAACaCCATGAAACTTGCTTTAAAAGTACTGAAATAGAGCTATGCAGTCACTGCATTCTGAAGGTGTTCAAGTGGCCAGGCTGGATGGGCTCCTTAAGGTCCCTCCACaCAAGGCATAcTATGATTCTATAATTGTCAATGTAAT

#106

GCCGTGCAGAGCTACAAAGGCATTTATGCTCTGCAACACTTTAGTCTCAAAGACTTCGAGGTCTATAAATACCTTAAACGCCGGAGATAAGCATCTGCCTGTGGAGCAGAGCAGAGCACCCACGGTTATTATTGTGTTAACAGAGCGCTTTAGTGGTCAAACCTCGCTCTGTTGGAGAAAGAGTGCCTGAAACGGCAGGCATTGCTCTGACTGAAGGATTCAACCCCCAATGGGAGGAAAAGCAACATTTCATTTTAATGGAAAAAAAAAATGAAAGCCCTCCAAATGTTAAACTAGCTGCACACATCTCAGAGCTCCAAACATAAACGATCCTGGCACGTGAATGTAACACTCTGGCTCACAGATGCGCTCCAAATGATATAAATTACAGTCCCCATCACTCATGGAGACACATCGCTCTCCATTATGCCATCAGAGAGGGAGAGCAGAGCTCGCTCTTAGAAACCAAGCGTGCIAAGCAAGGCAGCAGCAGAGGGGCTGCTGGCACTGCTCGTGCTGAGCTCCACTCACAGAAGTGAAGATGGACCCAGCACAGCCATGTGATGACGTCAGGGTCAGAAATGCAGCCCTTGCTGCATCCCATGCCAGTCCATTAGGCATTCAATCGAGGATGGTAGGAACAGGTATGAGGAAGGTTGGTTCTTCATATCCCTCAAGGTGCATGGAGCAGTG

#108

AGGGAGGAAGCACTGAAACACATAAAAGAGTTATCTCAAGA

#109

gtCCGCCAGCCGGCAACTCTCTAAACGCCCTGNNAAGCCGACGCCGAATAGCAACCAA TTCCCCCGGTCTTAATAATGCCGCTGACCTGTGGGTTGCCACCATCCATCAGCTAA GCCCTTAATCGATCCAAGTTAGTCCGTTCCCCCGCCCTGACCGTTCGCTGGGTGTCGGGGTCGAGTGGCTGCCGCTATAACCGTCTGGAGTTGGACTCCCCAAGCGTCCATCCATAAGGCATATGGGGCAGGTCTTAAATCACTCTCATTAGATTGTCAGCTGACGTCATGCCAGC AACGGGGAGGACATGAGCGCTCCACTCTGCCATAGTGATCGGGGATGTAAGCCCTG GTCCTGACCGTATGCCAGTCTGTGATCAATTGGCTCAGAGGGAAA

#110

CTGCAGTGAGTGGAGCACCACAGG

#111

CTGCAGCTGGCAAGAAAAGACAGGGCTACTTGGTCTTAAGAATAAAAGATGAAGACT  
AGGCTAAGTGCCTCTTCAGGTACACCGGTGTTCATCTACAGTACGTACAATCAGCAAT  
ACTGGTAAAATGAGCTGATAAAATTGTCCTGGCAGCCTCTGCTCACAGACAACCTCCA  
CGCACCTGTAAGGTTGAGGCTGTTCTACGCCGGAGCATTAAATTGCGGATTCCCA  
CCAAGACCAAGAATACTTCAGACTATTGGCTTCTGCTGATGTTCTTGTCTTGTCTT  
AACAACTAAACCTGTAAGAAAACAGCAAGTGACATCAGCCCTGTGCAGTAGGTTG  
GAAAATGCACCCATCTGCACGGAGTTAATAACACATCTAAACGCCATTCAACCTCGA  
AAAGCATTAGCATTAAATTGCACTGCTCATCTTGAGTGCAGAAAAGTGTGGAAAGA  
GGTGCAGAGGCTTGTCTAGATTACATGTCAGTGCTTAAACATGATGTGAGAAAA  
TTATTTAATTGCTGTACCACTGGACTGAAATGGCTACTTGCTAGCTTGTACAA  
ATGGTGCAGAAGTCTCTTGTAGTCAGGGAGTTGTCAGACAGAAATATGTATGTACTTGA  
ATGCCTGGAAAGCCTCACCATGAAGTAAAATGAAGTCAAATAGCCCCAGTAAGGTAGCT  
TTGCAAAACACTATCAAGTAATGGCAAGTGTAAACCACAAGACTACATATGAAG

#112

CTGCAGCTATAGACAAGTGACATACGTATGTATGTTGGCTCTCATTTCTGACACTC  
TGAATGTCTGGAGTTGTCTAAATTCCAGTTAATTATGTATAAACATAGAACAGGGTG  
ATAAAATCAGACAGTAGAGGATTGATTCTGTCCAGACTCTACTTCATACAGCCATCTGA  
ATGAAAAGCTATAAAACAAATTAAACAGACCTCAGTAAACCACAGGGCTTTGC  
TGACATTCTAAGGGTGGTAATTACAAGAAAAATTCAAAGAC

#113

CTGCAGTAAGAAGAGAAAGAAGAAGCAATTACAATTTCATGAGAATTTCACATTCTA  
T

#116

CTGCAGCGAACTCTCAGCCTCTCCAAGCAGAGAAGGGCAGAGGTGCTGAGCGGACCA  
CCACTCCATTGCATTCCCACACTGACTGAAGTTCCACCCACACAAAACCATACTCTAG  
GTGAATACATTCCATTAA

#117

CTGCAGGTCTTAGAGGCAGTCAGTACTCAAATGTTGATACTGCTAGATTAGCTGGTA  
TTGCTTCAAATCCATTCAAGAAAGTGGATCTGGAAGTCAGAAAGCAGAGTTTACATAGTC  
TTTGTGTTATAATGCAATTCTCTGTTAATAGTATTGATCTACATTCAACACTTAGTA  
CCTGGCTGTAGAACCTCTCCAGAGTTCCACTCGCTCTACACAGCACTGGGCAAAAGC  
AGATTGCCAGTGAGTCATCTATTCCAGCTGTTGAAAAAGGGATACAATATAAAATGTG  
TTTGTGCTAAAGTACCTGGATTGTCATCTCAGGTCACTACACAGCAGTTACTGAGAA  
GAGCCAGGAGAACTGCAAAGCACCTCCAGCACTCAGAAGCAGCACAGCACATGTGATT  
TGCAATCTGCTCCATGCTCTGAGGGAGAGGTTCTTATTGTTCTGGTCAAGAAAGT  
GTGTGTACTGTTAAAAAATAGTTAGCCTATCAAAGTAAACAGAGCAGTAGAGATAGT  
TGGCATAGTTACTCTGTTAATCTGCTGGGTGGCAGCAG

#118

CTGCAGTGAGGCATGCTCACCGGTACACAGACTGCTGCAAAACATCTGACAGCAAATA  
TGGTGGAAATTGCTGTGAGCTTATCAGGTACTTTCTGAGCTCCCTGCGTTAAGGTAA  
ACTTTCTCAGCAGCTCTGGAGATGAACCCAAGTTAGAAATTGGTAGTGTACTG  
TGTTGTTG

#119

CTGCAGGCAAGCTTTGAGCTCCTTTGCTGAGCTTCTCATTTAAATACAGTC  
TCGATGCTAGTGTCTGTTGTTCCACCCATCTCTGGAATGTGGGAAAAATTCA  
TTTCCATCTGGCACTAGCAACCATTAAATTCAACTATTGCTTACTCTCAGCTACTGGG  
AGCTGTTAAAGGCTCTCATTTGACCCCTTATGGCCTAAACTCT

#120

CTGCAGATCAAGAGCAGGGAGG

#121

CTGCAGT GAGTTGTGAAGGGAGGTG

#122

CTGCAGTATAAAGGGGAGATAACACTATGAAAAACATAATGTCTTGATACTTATT

#123

CTGCAGCCCTGCTCAGCCTGTCAGCTCTCGGGTGGTTGTGTGTGCTCCTGTCTCTCAT  
TCACCAGCTGGTGGGCTCTGGATAAGTTGAGCCAACGATATTCGTGTATAGTGTCT  
GCACAGGATCTTGCCCTCCCTTGGCAGAGGGGAGAGTTGGAAATTGAAGCTCTCCTA  
TCTCGTATTCCACTTGCCCTCACGTGAAGGGTAGTTGGTCTTCTCCTCATCCTGCCT  
GTTCTTGTGGATTCAAGGAAATCTGGCCTGGAACCTGCAGCTGCTGTATTGCACA  
GGTACCCAGGTAAAGGTGTG

#124

TGCAGCCCCGCTGGCTTGGCTGACCCCTCCCTGTTAACACAATGGCTGCTCCCTGCGA  
GCTTCTACGTTTGTAAATCCTGGGGCGGGGTGATTAAACGTTGGCGTATTCCAGCGG  
TTCCCTGTTGGTTTCAGTGGGGCTGTGGGCAGTTATTGTCGCGGGGGCTTGTGT  
ATTACTTATAATATCATGTACGGGTTCTGTACGGGTGTTCCCCTCAAAAAAAAGCT  
CCAAGAAAAAGTTGGGGAGGGAGTTGCGCCCATCTGGGTGGATGCTGTATCCTGTATG  
TTGAATAAGCACTGCTGATTACCTTACCGTTGCTGGCACAGGTTTACTCTCAAGTAACCT  
GGTTCGGGGAGCTCACGTTGGCAGGTTGCTGATGTTCCCCACTGACTCCTACGC  
AATTCAAGCGAGTAAAGGGATAATTGAAGAGGACAGGTAGGCTGGAAGGTGGCACTGG  
GCACCTGTCACTCAGCTGAGTTAATCTGGAGTGAACCTAGCCAAAAGTAAGCAAGAT  
AATAGCATCCCAGTACTTAAAGGGACTTATATATTAGATGGAGAATGACTTTACGTG  
GTCTGATAGT GATTGGACAAGAGGGATGGTTAAATTAAAAGAGGGAGATTGAGT  
TAGGTGTTGGAAGGAAGTTTTACGGGGA

#125

CTGCAGTTAAGACTGCAAAGAATGAAACCAATAGAACTGCTAACTCTGTCGTTGCAACGT  
TGCAGTGTGTCACAAAAAAGGCAGAAAAGCATTGAACTATTAAATATTCTGTTGAA  
ATGTGTAGCTAGTG

#126

GGGCAAACACAGATACAGTTAGAAGGGGCCTCGTCTGACAAAGGTGGCACCTGGAT  
TTCGGTGACCTACGAGCACAGGTCTGAGTTCTATCAGTCTTCTCCCCACTAGTGGG  
GTAAGAGAAACACAAAGGGCTACAGCAGATCCGCTTCAGCTGGTTATTCTGAGAC  
TAGAGGCACACCTACGGACAGTCTGGCTCTGGTTATGCCGACATCACACCTTGATGCG  
CTCACCATGATTACCAACCAGGGTCATGCTAATATAAGATCTGCTGAATGATACTCCAAA  
AATCCTGGGAAGGGAGCaaaaAAATAAAATCCACGAAGCAACAACACCAATGAA  
CTCGGTACTCTCATGGAATCATGTTACCAATGAAGCCTGCTTCAAGAACAGATT  
TGGGACCTGATTTCTGCTGTTCAATTAAATAAAAGTCAGTCTCTGTGATCTTATGGTC  
TTAAAGTGTGCAAGGAGTTAATTATTGCTGCAAATGCTCAGAAGGAAATGACCTA  
ACCGTAAGAGAGACGTGCTAAGAGCTCACATTAAAGCTCTGCCCTTATAGCTTGTCTCCC  
GCAAGCCTGGGCACCTAGGAAGCAAAGTAACACTAACGAAACACAAAATAACCTCCA  
ACAAGCAGGCAGTAGAAAATCTTGTAAACACATATCGAATTCAAGGGCAACAAAGCT  
TGGTCCATTACCTGACCATTCTCATCCATACGGGAGATTCTCAGCTAGATTCTGCTAC  
AGCTCAAAGCCAGACTACGCCCTGCGAGAGGGACAACACTGTGAGGTCTGCACTTCA  
ACCTTCTGCCATTACTGTACAATCAAATATGATTGTGCAAGAACATT

#127

CTGCAGGAAATGAAGAGGCTGGAGTGCTCTTTCTCCACAGACAGAGAGAAAATCT  
GGGCTACATGGACTGTTGGCTGACCAAGCAGGGCGTTCTTATGTTAACGATTTGTTA  
CAGTTGAATTAAACATCAGTGGTGTAGGAGACAGTAGTGTAGGTTATCACACTTCTAA  
GTTTGAGACAATCAAATATGTAGCATTAAAGAATACCAAGAACCTTATATTCAATCT  
TTGGATTTCTGCTGGCCAAACAGTTAGAAATTCAATCCATAAAAAGTAGCTTT  
GATTATTGAGCTACAAATCATGCAAGTTGTTAAAGAAAGTTCCATACAGCCTTGTAT  
TGTACAGAAGTGAAGATAGATTATTCTGCAATAATTAAACTCATATCTGATTATTTA

TCAGATTATGCTTCAGCAATACTCACGTAGCAACTTCAAGAACTGAACCTAGAATT  
CTTGAAAAAAATGGCTCTGGATTACACTGCTCCAATTCCAGCAAATAGAACATATACA  
ATAAACCACTTCAAATAACAACAGACAGGTACTACTGACAGTAGTTAAATATTGCTGAT  
GCTCCTTGTATCATAAGGATTCAATATTGGTCTAGGTGCACGTCAATTCTACACCT  
ACACTGCTCAAACATAGTTTATAATAAAATACCACAGTCTTGACAGTACAGTGGGAT  
TGTTTATTACTTACCATATAAAAGGAATATATTCCCCTCAGTGAAGACCTACTTGGCTT  
TTAACATTGCTAATTGAAAGCAAATCTGAATTGCCAGAGCAACTCAGGTGTCCTGG  
GACCAGTGAAGTATAAACTAACTTATAAAAGAGAGACAGCTAAACTATTACATTCT  
CCCTCTTGGTCAAACAGCACCACACACACTCTGGTCTTCACTGCCACCATAGAT  
AAACGTTAACTATCTAACCCCTATTCTACTCCTGCTAAAACAGAAGAGAAATAGCACA  
ACAGTTGCTGATCTGAATCATCAAAAAAGATCTTTCTTGCAGCTACTCACTTGT  
CTAGGAGCTGAGCAAGTCAGAATTAGTATGAGTCAAAGGAAAGTTCTGAGAAGTATT  
TGAGCCTACTTCAGTAAACATTGCTGAACTATCTGTTGTAATGCTGGTGTAAATC  
CTAGTGAAGGCAAACAATTATGTAAGCATTGCTTAATTTCATGTTGAAGATTCTTC  
ACACTGAGGCAAGTTGAGACAAATGTGTACAAAAGATTCAAATTACTCTTAGACT  
ATCCTATTITGCAGAAAGTGGAGGGTAAAGCTAAATTCTGTAACCAGTTAGAAATAAC  
TATTAGACAGCCACTAGATCATCCTCTAACAGGATGAATAGTTGAATTGCCACACT  
AATTTAACTGAAGTATTGGTCCACTAACAGTGAACAGCTGACTACAGCTGATTGTCAGATAT  
AGCATCATGTAGAGTACTGTAAAGCTGACCATGAAGTAGACCAGGATACTAATGCATA  
TTAGCCCTGCTACAAGTTACACAGGCTTCAAATAAGGTAAGGTGATAAA

#128

CTGCAGTGTGTTATAACAAGGCAGTTCTCAGCTCAGGGGCTTGTGATCACAG  
CAACAGGAGTGGATTCTTCTTATCTCGGATAAGCAGCGTGGACTCTATTAAATCAG  
AAAGGGCAACGCACTGAAGTCTTTGTGTTGTTGTTAAATTAAAGAATATATATTAAATT  
CTGAATTACATCGTGTGTATGTTGTTGTTAAATTAAAGAATATATATTAAATT  
ACATATGAAAGTCAGTTTATATATTCAAGACTGATGTCAGCTGAGATTAAATT  
TTTTTTCTGGATAATGTTGACAGTTGAAAATCAACCACTGCTTTCCCCAG  
TATTGTCAGTATTGTCAGGAGTAAATTGTCAGT

#129

CTGCAGATGGAACTCACAAGGGCAGAGCAGACGGGACAATCACCTCCCTTCCAGCTG  
GCCACCCCTCTGTTGATGCTGACCTTCCGACTGCATGAGCATACTGGAAGCTCACGTCC  
AGCTTTGTCACCAGAACCCCCAAGTCCTCTCACAGGGTGTCTCAGCAAGTCTT  
CTCCAGTCGGTACTCATGTCAGGACT

#130

GCAGGCCTCAGCTCACCTCACGTCTACCGTGGCTGGTGTGCTTAGGTCTTGGTAA  
ACAGAAAGGACTGCTGATGGCGTGTGGTGGTTGATTGTTGTCGTGTTTTAAAC  
AACTGGAAATGAATGAT

#131

AgACCTATAGCTGGGGATTGGCACCACTTCACTTTGAAAGGGGTAAAATCTCCT  
CGTTCTAAGGGATAAGCAACTTCTCAAAGGTTAATGCTGAAAACGGTCAGTCTGATT  
TTGGGGATAACATTATTCAAGTAAATATAACTAAGGTTCTTATTAAACCTAGGCCACT  
GTTAGGAGAAGAGCGGAGGCTGTAGTGTGACAGATGGAGTAAAACAGCTTATTGG  
GGATGCTAGCAGGGAAAATGTAATATGCTTCTTATTACCAAAGCGCACATCTGAAAAC  
ACAGGGGCTCATCAGAGGGTATGAGATACTCGAGATATTCAATGAGAGGGCAGTGAG  
ATGCCCATCACAGGGCCTAGGTGACAGGGCTGTGACTGTGAGAGCTAATGTG  
TGCAGACACCAATATAGTGGCGTTGTATATCCTAAAAGGCTGTTAACACATTCAA  
TCGCTATCCGGAAATGCTGGCACAGTGTGGTGTGACACCACCACTTCATCTGTC  
TCTTGCTGCTTCTGATAGTCTCAACATTATTAATACATTAATCTCAATGTTCTG  
CCTGGAGTTCTTTATAGAGACCTCTG

#132

AGAGGTAGGGATGGATGTCTTCTCTTGTGACCTGTGACAGGCAGTGCAGGATGGTGTG

# 134

GctGtAAgCACTTGatatCaCAGAgATGGaGCTcATacCaaCACCTGtGCAgCagAagAaATATGtgT  
GaAAGAtAaaACTGgACCTTGCagGAaAGcCTtGAAcAgAaaTaCaAAGGtaaAACaTCCaAATAAA  
CAACTACCaacTGAAcCTTGTGGGTTtTTtCCCaagGtaGtTTTTTCagctTtgatacaCCtGTTccT  
AAGtacagtcaATTtaCaC

# 135

CTGCAGAGCTGACATTCTCTGACATGTGAGACATTAATTAGGCTTTGGGAATGTCC  
AGGCAGAAAActGGTGATGTCAGTCAGTCCTCGTCCCCTAGCACTCATGCCATAGTTGAAT  
TTAATCAGTCCTGTAACCACATCCTAGTGCTTCAGTACCTGAAGATAATCTTCCACTT  
GCTCTCC

# 136

CTGCAGAGTAAGACAATCACTGTAAAAATCAGACTGCTTGGCATACAAACAGCACT  
TTACCAAGGAAAGGAAATGCAGCTGGAAAActTTTTTTTT

# 137

CTGCAGAGCCCAGAACAGTTGCTCTGAACACTGATTGTGGAGAGCTTATCACTCTCCTCAC  
TGTTCACAGTAAGCTGGG

# 138

CTGCAGTAATCGTTTCACTCAGTATACTGTC

# 139

CTGCAGTCTCTAGCCCCAGTCGTTTGCCCTGTGCTAAAGAGACTCCTGTTTATACTCTT  
GTCAACCATAAGACCCGACAATGTAGATTTGTCATGATCAAAGTTTCATTACTTTGGG  
AGAATCAGCAGTCAGATAGATAAGCTTATCCTTGGAAAGATGTCGTGAGCACTGGTC  
TGTCAACCGTGATAAAGCAGCTGTCCCAGCCTCTATAATGTTGATAAAACTCCTTATGA  
TACAGACTGTCATCTTGAATTACAGAAGTGAATGTGGAATGGATCCACCGATCTCGA  
TTGCAACCTTCACTTAATACTATTGTCACTGCATTGCTACTTCTCTGACAGACATAT  
ACCTTCGTACGACATATCAAACACTAGAGGCTGGCAAAGATCATAGACTGAGCAACT  
CTCCAGTTGATGTTTATCCATAACACTGCTCCACAAACAAACAAATGGATTCTCTGG  
GCTCTTGAGTCTCTGAAGCTCTCTGTGTTCCAGCCTCCTTCCCTTCTCATCCATC  
TTCTTTGCTATTTCTGTACAGTCTTAAGATGCAAATATTTAAATACTTCTTTT  
GGCGGATTGGAAAGGACATTCCATAAAAGCTTTAGCTGTTCACTGATGTTTCAGG  
AACTGGCCGACCAACTAGCCTCCACAAACTCCAAAGTCTCACGTG

# 140

CTGCAGCAAGCTGAGGCCAGAAGTGCCTTTA

# 141

CTGCAGCCATATATGTCCTCTAGGATTCCCTATGGCTTTCCATGGCCAAGCCTCCATGAT  
TAGTAAGAAGACTGAAGGCCATAAGATGG

# 142

CATGGAGGTCTTGAACAGGTCCAGAGGAGGACACTAACAGATGATCAGAGGGCTGGAGC  
ACCTCTCTATGAGGAAGGTTGAGGAACCTGGCTTGTAGCTGGAGAAGGCTCCAG  
GGAGACCTCATTGTCCTCCAGTACTTGAAAGGGAGCATATAAACAGGAGGGAGAATG  
GCTGTTACAAGGGTGGATAGTGTAGGACAAAGTTgAATGGTCTTAAACTGAGACAGGT  
GAGGTTAGGTAGGTAAAGGAGGTTGAGTGGATGCCCATCCCTGGAGGCATTCAAGGCCAGGCTG  
ACAGGTTGCCAAGGAGGTTGAGTGGATGCCCATCCCTGGAGGCATTCAAGGCCAGGCTG  
GATGTGGCTCTGGCAGCCTGGTCTGGTGGTGGTACCTGCACAGAGCAGGGGGTT  
GAAACTTGATGATCACTGTGGCTTCAACCCAGACCATTCTGTAAGGAAATAAGGTTGCTG  
ATTACAACTCGCTGTGTTGTACAACACGTTGATTTGTGAATAAGTAAGGTGCTG  
TCTTTGGTCAGTTCTCAGTAGAGAGGATATTCAGGAAAAACCTTCATGATGGAA  
TATTGATGAACTGCCCTTTGCTTAGtGAA

#143

CTTGGTTTTCTTCTTCTGCTTTGTTGGCTCCGGCTCCTGCTGACTCCCTTGCTC  
ACGTTCTGGCTTGTGTTCTGAGATGGAATGTGTTTGATGGTATTAATAAGAAAT

CTTTATTGCACCAGCAAGAGGACATTCAACCTGAAAAGATATAAAAGCAATCAA  
AGACAAAATAAAACATCCTGAGATGAAACTGAGAGTTAACCTAATAATTAAAGAC  
ACGTTCGTGTAGCTTCCTGCATGCCACCTAGTGAAACTCTCACAGCACAGAAAGAT  
ACTCTCAAGGGTGTCTCCATTCTCCTTACCTTACCTTGTAGAATAAAAGTTATTACTT  
CAATGCCATGAAAACCCATGTGATTTCACAAGGCAGCAGGAAGCAGAAGGGCAGTGC  
TCATGTTACCGCACTCTTGTAAAGCACTGACAAACAATTATGTTGTCACCTAACCTAGC  
TATGCAGGAATCAAATACTTAATAACATTCAAGTCAAATGAGAATGTTTACAAGACCA  
TAACGTTGACCGTAGTAGATTAAATAACTGAAATATCATACATGAAATATTAAATATAG  
AAAGTAAAAGAACATTGCATGAAGAATCCAAGCAACTGCAG

#144

CTGCAGAAAGACATACCCAACCGCTAGTC

# 145

CTGcAGAAAGCTAATAGCAACCATTATGAATCCAAATAAGCTGAACCTAGAAGAAAT  
GAGACTGTGCGGGCACAGCATGCTGAACAACAAGGCTAACGTGACTGACTGCCTCT  
TCAGCTCAGTAGGCTGTGAGGGAGTATGAGTTGTGAATTGACACTTGCTGATGTA  
TCCCCAGTGGAGTAGCTGAACCTTATGATGCCTGAGTTCTTGAAGCATTAGGGACT  
GGTAGTCACAGTGTAgCGGGGCTTGAGAAATGAGTAGATATAAgGCAgGTCCATGCATG  
ACAGCTTCTCCCCTCCgCTTCCCTCCTCCCCAGCAAACCAACTCTCACCCCTGGT  
ATTCAAGACTCTTTCTTATAGGTGTTAGAACAAACTACTCCAACCTGTAAATAACAA  
ACTTCACAGTTTCCCTTGTGACATCCTGGTAGGTGAAGCAGAGAAGTGTCTGTGG  
CACATCTAACAGCAGATGAGAGCTAACGAGTTCTTAGCGAACATCAAGGGATTAAATCTGT  
TTGGGGTAGTCTGCCACACATCCTACTTGTGCACTTACTGCAACTGCAACTGTTCTGT  
ATTTGCTGAACGTGTTC

# 147

CTATCCCATTGCATGTGATACTATTCCCTCTCAGCTGGAAATACTTCTGTATCTTGC  
TTGCTTCCAGTTACTGCATTCACTGCTGAGAGGGCAAGGAAAAACCCAAAAGTGC  
AGTAATCATCTAACATAATAATCAAGGAATAGGCTTTGTATTCACGAGAAGA  
AGCATTCTGCATGCACTTGCAGGAATCTGTACTTCTCAGACAAAATAAGGAGCTT  
AATTATCTTCTGCACAGGTCTGTCAAATCAAATCACTCTAACAGAACCAACTAGGC  
ATTTACATATCTCATCTGCAG

# 148

ACATGATATTACTCTTATGTAGTACATCTTTAATGTCTGGCTAACATTAGATC  
CTCTAGAGATTCAGTGGAGAGTATCAACCATTACTCATAAAATCATCTATTGACATG  
GAAGACCACACTAGGTCAAGGTCAAATACATCTGTTCTCTGTTAGGTGTAAAGAAAGC  
GTGAGGTGACTAGCTTATGTAGATGTCAAATTGCAAGTAAATGATGTCAGATATATC  
CTCTCTCTGGAATTGGGGTTAGAGCTGTAAATGTTactagctggtgtnacagcagcattaccactgtattgt  
ctgaggagagaggaatacaggaggtagctgtatattcatataatatgttatcttggaaaggaatatgtatggaaaatagaaggatatcat  
actggctgcag

# 149

CTGCAGtGCTgtTttGtCAgTtGGAGAGGCTTCANTaGGAATTAgGGGTGAACACT  
CAgCACCTGtAAGAaggCTCTCaGCGCTgCaCAACGGGG

# 150

CTGCAgAtgAAACCTATTGCTGGTTGCCTAGAGCGAAGTAGTCTTAATTACTGAACTA  
TGAATCTCCTCCCCAAAGTCATATGCTACGTGACCCAGGAGGTGGGATCC

# 151

AGGTACGCAGTCAGGTAGCCAATAggAcTTTTCTCCTGTCATGCAGAACGTTGATTG  
TTAACGCCTGAAATGAAAACAGCAGAAGGTGTGACTCTACTGACACCTGGAATAGAAAG  
CTGTAAAAAAATTACTCTAGCCTACAAGTGACAGTAAGCTAAAAAAGGCAAAGCTGTT  
ACTGCTATATCAACAAATTACACAACACTCAGACTGTGATGTTAATCAGTCTCCATCCA  
GAGAATTAAACACGAAATCCCTTAqAAACAGAAATATCTGCTGTGAAAGGACATGTATA  
TGGAAATGTGAATACCAAAGAAAGAAATGAACCTACTCTCAGAAATTAGGATTACATCT  
GATAAAAGTCCGAACTTGGACCAAATCCTGTTAACGCTATCTCACTCGACACACAAA  
AATAATTATCACTCCAAAGGAATGGGTGTAGTACCAAAGGAACCATTTCATAAGGT

CAGAGCAGAAAACGTAAATTATTTATTATATAACCACAGGCAGTCTATAAAGTAGTA  
TTTATTTACTTACCATAGTATTAATGCTTATGAGCTCACTCCCACGGGATGG  
TTGTTAGTCATCAGAACATGCAAATCGTCTCATCATCAGGGTACAGTCAGAAC  
CTGGCATTGACATGTTCTAGTGCGCTCGGGTAACAAATGTGGAAGTCAACAGACACATA  
GAACACGCCTCCTCCTCATCACTGTGTTAGTGATAGGACTGTCTAAAGGAgAAC  
ACTTGAAGGtTTAACTGCTGAACCTTTTATTCTCATCTGtGtTCTCAGCAaCTGT  
TTGcCTGAACTCATTTAAGACAAAGCGCAAGGATGAGATACaaTGcTTaAGTTAGTATTTAAA  
GAfTTCAAGAAGGAAACAAACAAACTATTatGtaCaCCttAGTCTCAAAtag

# 152

CTGCAGAAACATGCGAACGTGAGGTATTGGGAAAAACAACAAACAAAAACAGTGA  
GGAAAACCCAACCCATGGCAAGAACCAAGGATTCTCATAAAGCATACACGGGT

# 154

cTgCAGTGTGGCAGAACATGCGAGCAGAgTGGTAGATAGATCAGAAGTGGAGCTAGACTGAA  
TGTAGTATTGCTTCATCACAGACAGGGTCCAGCCTTGAGATTGCACTGCTGATACTAAAG  
AAAGTAGtGTTCTGCTGCTGATCTCTTAGATCAGATCTCACCTTAGTCTCGTTATTT  
TTCTAGGATCTGCTATACTGGGCTCCTGtGGCTTCCATTGGAGITTGCAATTAGC  
GTAATTTAAAAGGCCAGAAATCTACATTCTAAAGTAGTCAGTGTCTGTTCTAAG  
ATAACTGTTCCATACAC

# 155

aTGAACgCCggaaTTAACTCGagagGGATATCGaGtGGCcGCCCGTgCAGAGCGGCGAgA  
GCTGCCGGGGCAGCCGGAgCTGCCgCGGGGaTGCgATGGGCGTTGCaGCGCACGGACa  
GCAGTGTGGGTGCGATGAGGAGCAGGTAGCTCCAAGCGTGCCACCGCTCAgCgCagTTGgC  
aTTTGTGCCAgcTGGCCgtgCCGAGCCCGCaGCAGCGTGCCTACAGGgaAgAGAgATGCGC  
TCGCAcgTaACAAAGGcGCTTGGcTGATCtCCCGTTGAGgAAcTGAfTCTTCAAAATCCG  
CCTTCTGGcGTTCCCCGCTGCCTCCGCACCCCTCCTTTCCGGtGGTTGCCGcGCCGGGCT  
GGGCaCTGTGCAGACGCAgCAACCGCTGATGCTCATCCCTCTAAATCCGCTGCC  
CGCTAACCGCCTCgCCAAAGCCTTGACACGAGGCGCTaAGGACCGGCGTAgCATTCCaAG  
CCTGCGTTAGGTTGGAGCATCCCTGGCGACGGTATCGGGAAATAAAACCCGTGGCTTG  
AGGGtAGATCCCAACGGtGCGAGGGAGAAGCGGCCAGAAAtAGGCACCTGGGCGTTGCT  
GTGCAgCeCaGCTGTGCGCGggGAGCAGcGAGCGGAGGTGGAGGGCCAGCTCCCTTGgCg  
TAgCGGCGGGTTCTTATTgTTAGCgTggAAAGACGGcCAAAGATAagGAGCGGTTGTCATCC  
GGCTGCCGGGAGGACTCGcgGAACAACAACAAAGAGCTGTGtGGAGGtGGAGATgtTTTC  
gTA

# 156

GCTGGCACAAATACAACAAAGCAAATTATTGCTACCTGATGATTCTACCTGCATTCCCT  
TTTCAGCTTCTTGCATGCTTACCTTTCAAGAAATACAGGCTATAAAATGTAACCTGTC  
ACACTGACTGTGCATTAAGAACAAACTCAGATATCTCTCTTTCTTTAAAGTGTAAA  
AACCAAGGGAAATACACCAGTGAGACTGCTCATGAGAACAGATCTAGTTCTCATGCATC  
ACTTATTGTTCATCTGCTGTTACTGCAAGTTAACATAGGACCAAGAAACCA  
GGAAATTCAAGAAAGCTTCGGCATTAATTAACCAT

# 157

CTGCAGGGCCATTCCATCTTCCCCCAGAGCCCACCTGTGCACTCAGAGGGAGCTGC  
TGAGGTCCGAGCTGGCTGCTCACCTGTTGAACCGAACCCAGGGCACCAACACGGC  
AACACACACCTGTGGGG

# 158

GGTGTATCTCTACAGCAAGGCTTGTAGATGTTCCAAGAGGTCTGGCTGAATCCATA  
AACTCCCACTCATCTCCTGAGTATCTGTGGCTTTCATACTTCTCATTATTGTTCTGG  
TCCATTAGAGTTCTCCACAGATTACTCCCTCAGTGTGTTAATTAAACAAAACATAAT  
TTGGCTAATACTGGATTGTTCAATCCAAGTAAACCAAGCTAAATAAAAGCTG  
AAGTTCTGGCAACACATTCAAGTCAGCCATTCACTAGATAAATTGtgTcTcAAAGACTGG  
ATCACAGTGCATAATaTattaaaattattggagc

# 160

CTGCAGTGGTCCTGCATGGATCCACACAAGTCTGCTGGCGTTATCTCTCAGGCACITCA  
TCTGTTCCCTCTGTGCCTTAAAAGGAGGGTCTTCTGAGGGTCTGAGAGGTTGCAGCAGCA  
CGGAAGGATCCCACGGTGCTGCCACTGCAGGCCGACCCGAGGGCGGGATATTCTG  
CTCTCCTTCGGGCTCTGTGTGACACTGCAGGTGCCACCACCGTGTACATCGCGCTTCA  
CGCTGCCAGCCACTGTCTGACTTGGCTCTGGCAGTGATGTATGAACTCAAACCTTTA  
ACCTGTGTAAGTTATGTAGGTTATAAGGCTGTACCTCATTGTGAACTGCCAGCTTTAA  
GG

# 161

CTGCCACAGGTAAATTGCAGAACTGTATTACAATTCAAGTCAGTGGTTCCCAAATCA  
ATTAATTCTGACTTACGATTTTAGACATGGTCTGCATTAGTCTGCAATATGATTAAAA  
AAAAAAAGtGCCATTGCTAAACACTTAAACATAACTCCTCTAATCTAACAAATACCCCTA  
ATTATCTTCTCCCTTGTTCAGTTATGAAATAAGGtAGGTAGCAATGACTCAgAGTAA  
GCACTGAAAACACATCAATGTGTTGCCCTGCTGATTCTAGGACTTAataagt

# 162

cTGCAGcTCATTCATCATTAAgCAACTAAAAGTGAATgCAgAAAAGACAGGGTGTGAAGG  
TATTGAACCTTTGTGCCCTTATTGAGGTAAAACAAAGTTAATAACTGAAATGGACAGC  
TATGAGAGGCATTTGAAAACACTGTTTATGTAATAATGCATTGCTTACATTATTTGTA  
CAAATCCTGTACAGACaGCCTTGGGATGGTTTGATCTAGcAAGGAGcTTGCAGATT  
GCACAGAGTTCTCAGAAACTGCTGTGACACCACAGTCCCACACTACTGGTTGCTGTCTTA  
CTGTGGTATTGTCGGCAGAATTCTAATGACACTGCCAACACTTGGGGTGGAAAATCAA  
GAGAAATAACTCTCATTTACAACCTCAAGAGATGCATTAAAATGACAATGTGGTCTGTA  
TTTAAATTAAACCTAACCTTACCAAGAGATATAAC

# 163

GGCACAGATGCTGTTAATACAAAGAAGGCACAGATTGAGAAGTTGGCTTATTCAA  
CAGAGCCCTCTGAAATGCTGAGAACCAAGcTATGCCATGGACTGGCCACAGCACGGG  
ATCAGATTAGAAAACACaCAGATTCAgAAAAagtt

# 164

CTGCAGAACGGaTGCCTTCCAGCAAGTAAAGCCTGGTAGTCTGAGAAAACtGAAGATGG  
AATATAATAAAATAACAAGCATGTGAGCCTGATACAAAGCCTAGTCATAACTAAAGTCA  
AGATACATCTAAAAAAAGAACTGCaTTGGCTTGGCTGAGCTATTACAATTCACTGTTGA  
TCAGACTGAAGTCAGTACAGTTATCTGTTaaaaAGGTAGCATACTCCTTGCAAAGA  
CAACACTGTAGGTGTATAACCTACTAATGGATGTAATGGCAGAAGCAGGAAAAAAAG  
AAAAAAAGAAAAAAAGCTACACATTATGTAATTATTAACAATATGTACCTAAAATCTCTT  
GTTCAAGtACAAAATAAATTCCACCCATTATAAAAAAAATCTAAACCAAAAGAAAAAGA  
AGAAATAAAAAAAAGCAAAAGTTAGCAACACTTTATTGGAAATGCAATGGAAAAAGT  
CAAGAATTAAAGCAAAACTCC

# 165

GCTTGTCCCTGCTGAGCTCTGCTTCCAGGAGCAAAAAACATTAAAAAAATATTGAGA  
GCAGCAACATCAAAGAAAGCcTGTGTTGGACGTGGCCAGCAAGGAGAAAGCAAATAAA  
GAGGGTAGATTAAACTAGATATTAGGAAGAAGTTCTTACTGtGAGGgTGTTGACATAC  
TGgAACAGGCTGCCAggtAggCTGCGgATGTGtGGATCCCCCTTCTGGAAAGCACTGAAGG  
CCAGGCTGGATGGGGCTTGAGCAACCTGGTCTAgAGGcAAGTGTCCCCGTCTaGCaGG  
GGtGGtGGAAGTAGGtGATcttaaaGcgTcccTTcCaaCccaaatcattctgtgatttatgaatgaaaataataat

# 166

GGTGAACCTTGGCTGAAACACACTCAGCAAAGAGCAGAGGGCTACAAGTAACCTCTT  
ACCTGCAAGAATTGCCTAGGAGTAGAAACATCCAAAATCCTGAGAAAATAAACGGTC  
TTAGTTTGCCTCTTTGTTCATGCATATTCCCTTGCTTCTTACTAGGATCAAT  
CTAGTGCCTCATAGGACTCCGCTCATCACTTCAGCTTGCATAGCCCTTGTGAGATG  
GGAATGGCTTCTGCAAGAGCAACGGCCAAAGACCATTGTTGGAGGGTGGCATCCACCC  
AAAAAGAAAAGGAGACCGGGGCTACTGCTAGAGCAGACGTGGAAAATCAACACCTTG  
GCAACATGAGCACACTGCAG

# 170

TAATTTATTTCTAAAGTAGTTGTTgtAgAAAAgAAAGGAACACAAGTGCAGTT  
TATTGATATTGGCTAATAGCAAAGCAAGACTAAAATAAGCCaCACAACCAGCAAGTGT  
GAGCTGACAAGCAGCAAGCGACAGGATAGAAGTGTGAGAAGTTGGCTTATCAGT  
TAGACAGCCTGTTGGGATCATCCGAGTCATTAGTCAGTGCTCCCTGGGTTGTCGCTGTTA  
AATAGGGGAAGTTAGGAACAAATAGAACAGCATTAAACAGCAAAGAAAGCCAACATG  
TAGGCATTTGACAACAAAGCAGAAACTACATAGCCATTCTGATGAAACTCATTATGCCA  
AGCACTTGCATAGACTTAGTATGTTCTGCAG

# 171  
CtGCAgCTGCTGGCGGTGTCAAGCGTGGGGTGAGAGCTGCTGTGCAGTGGGCCACGGCCA  
GGGAGCAGATGGGGGGG

# 172  
CATGCTGGACAGGAAAGGACAACAAGCAAGCAAGCCAGGGCATGTGATGGCAATGCA  
GGAGGTGGACAGGGACAGATGCTGATGGAGAGCTCTCCTTAAATACAACCTCCCCAGC  
TCAGGCCACTCACTTATTCTGAAGGCCAGCAGTGGTTGTTCAAATCGGTCAAAGT  
CCATCATCTAAATCCCAGCAGCTGCTCCCTGTTCTGAGTCATCTGCAAGC  
ATTGTCTGGTCAAAATAGCACTCGGCACAGTCTCCCTAAAGCTACAGCTGCATCCCT  
ATTGATCTGGAtGtctGGCAgAGCAAAGGAtggCCTTCTCCCTGCCTTCCCTGCAG

# 173  
CTGCAGCTAGGTTCCAAGCAGAATTAGTACACAGAGCTGTGCAGTTAATTAACCAAC  
ACATTCTCTGGAATATCAGCAGATGGCCACTGAGACGAGCAAAGTGCCTCATCTTT  
TCTGTGTGCATCAGTTCTCCGTTAGAAGAAAGGGCAGGGAAAAACAAAAATGAGTTT  
GCTATTAGACGTGTTGGAGTTAGCTGCCTGGAG

# 174  
CtGCAgCACCCTGCACTGCTCATGTGTTACTCGGGTATGTAGTGCAGTCATGGAAAAATA  
AACCACACAGCAAAAGGCAGCACGAGCAGAGAGACAACACGGAGAGAGGGCAGGGCA  
GGATGAGAAGGAGTCCTGCAAGCCATGTGTTAAACCAAGGGTGAATCAAACAGCGC  
CTGTCTGAGCTCCCACACCTGCAAAGGTCCATCTGTTACACAGCTCAGCCGCTCTCCTG  
CACACAGCGCTGCGAACGATGCCCAAAGACTGAAGGCATTCTCAGCACTGTGCCA  
TGGGATCCCAGAGCCATCACCACCTGCTGCTCCACAGCGCGCGAAGATGGAAGTTCT  
CTCACGGAGCGGTATTACGAAAGGAGAAGCTTTGTTCTCCCGAACGACAGTCTGC  
ATCAACTTCACCTCAGCCGCGCGATGAGAGACAAAGCACGGCTGGCCAGCCGTGAG  
GCGGAGGGCACC

# 175  
CtgCagTTTGTAGGAGGGAGCCTATAACTTGGAAAGATATTGTTAACTCATAATGGACAAAT  
GACAGTGCAATTACTACTTCACATATTCTCTGAAGTCTGCTTAAGATCTCTTTGGGT  
TTAGAATCCTGTTAAAGGTCACTGAAAACAAAGCTGGTTACCGCCTAGTGCCTTTCA  
GCCTTCATGTCCTGAACATAAAAGGAAATTGGGCTCGTGGCACGTTCATAAATGATGTA  
TCTTCAAGTGAGATGACTATTCCAGTGTACTGCTACCACATGAGCCTGGCAGATTGC  
TTTGATACAAATGATAGTCAGAGATCTTCTTAAAGAAAACAAG

# 176  
CtgCagTTTGTGGTTGTTCTTTCTTCAAGTCAGCATTATTACTGCTATATGGGAGATG  
AAGCATCCTCATCACTCCAAACTCCACAGCACAAGAACGACCAACCTCTATTGAGAGC  
AAACCATAGGCTCAGCCGCTCAATTCCACTGCCAAAAGTTTCACTGGCACAAATGAAAC  
CTACTTGGCTGGTGCAAAGATCAAGATACATTAGGCATCTGCAAAGGAGAGGCAAAGG  
AAAGGTATTGATGGGCAGCAGGG

# 177  
CTGCAGCAGTGCAGGAATAATATAACTGAACAAACAAACAGAAGTACAAAACCCTTC  
GCATCC

# 178  
ACAGTGGATGATGTGGATGTGCTATATCCACCAAGAAATGAGGGAGCTGTGGACTTTatt  
TCCCCCAAAACAGGAGATAAAACTGCAG

# 179  
CTgCAgAATGCAGACTGTGGTTAGCTAGGACTTTGGGAAGAAATAGAAAATTGGtGGCC  
AGGAAATGCCTCAGAGAGACCAAGGAAAAGCCAGAGCTTCAGCCTACTTGGATCAGG

# 180  
GCTCGCTGGACATTTTATAATTCTTGtGCTGtGTTGcTGGAGATCTCCAATTAGCCTCAA  
ACCGGAGCTCGCAGCTTGtGtGCGAAACAGGCAGCTGCTGGCACGCAATCTCCCCGAAG  
TTGTTTCCCTTCTTCTGAGATTAGAGcGGGGGGGGGG

# 181  
CtgCAGCCAGCGGAGTGCAGGTAGTAGTGTAGCAGCTCTGAACACTACGCTCTGTGTA  
GTTGGGCACGAGGATGGG

# 182  
CtGCAGCGCAGGTCTTAAtgcTCCAGGAGaGCAGaAgAAAAACTTCTGTTATGAGGAAATG  
TGAACCTGCTTACTTATGTATGATTAGTGCCATTCTATTCAAGGGAGTGCACGAGAG  
GTCCAAATATTCGACTTTGTATTATAAGAACACTGCCATTCTGGCTCTTACAGAAAGTA  
ATTTATTCTCTCTCCCTACCTTGGGGAGAACATTATAACTCCTGGTTGCTGGAGA  
GTCAATAAAACAGCAGTGCTGCTAAAGATCAGCTTGATGCTGAGTGAAGCAGTGA  
TGGGGAAAGAAAGGGATCCCAAATTCTCCTCTTCTCCCTGCTATGTTACAT  
CCATGAGCGGAACCAAAACACAAGAATTGGccCttCTTTCTCCCTGCAACAAACGATTAA  
T

# 183  
AGCTTGATTCAAAGAACGCCCTTCCTTGgetgtAGcACTAACATTATCCAGtTCTC  
CCTTAagCTCTGAATGtTCAGctcAtACCTGcAAGGcAAGGAGGtGAtggcGcaCACAGATAAgA  
CATTAAGTGGCTCTCAGGGtCGGAaCTCAGtACCCTGGAGAGCTGAgCCAGAGTCACTTATat  
GAGGGataGGTCTTATAGCCCATTCCACTGCTTCTTGGAGTGAACATTGAATTGc  
AGtGCAGTGCTGTCACCaGCTTTACaATTGAGGATGtGTTCTTGCCaCTGcCTTTTTTT  
TTTT

# 184  
ATTATCTAAAGATGCAGACATAGAGAGAAACGtgTGCAAgaATGCCTATGAGAGATGCCGC  
AAAGCAGATCTTTAGGTTAAAAAAATGTTGCTGAAAAGGACTTAgctatTTCCTGCTTGA  
TGTTTACACCATAATAGATCTCCTGATACCATTATGCTGTAGAACCTAATGAGTTAC  
TTGGtGtaGGAGGCAAGACACTAAAACAGGcAAAGGAGGAGGAAGGttaTTATGCAgCCAG  
AGTGTtTACTATTCTTGCTGTTGTTTGAAaCAGCTaTTAAACTAGAGCGgCCcTCGGg  
aAAag

# 185  
CTGCAGTAAATCTACTGTAAGTGTAAACATGGTATGTCAGCATTAGTCAGCTTGTG  
TGCTTGAaaaaAGAAGTAGGTAGGAGATATTCTCCAGAACAGAGTCAGTTGAGCAGTTT  
GCTAGGTCTATGAGAATCTTCTCAGACACTCTGACTACACAAACAAGTAGTTGTCAGC  
CAGCTCAACTAAGAACGTTGGTAACATCTGTATGTCATATGTCCTTACCTGTATTAAA  
GACATTGCTGCTAATATCATACCTATGCATAAAACATTGGTAAAGTGTATTGAAT  
TTAGTTTATGGAAAATAATGAAACTTATAGGGTATAGAGATATGCTTTGCGT  
ATACGGCGT

# 186  
CTGCAGGCTGGCCTCCAGGCTGTGGATGCTTGTGGAGCTCCACCTGTGTGCAACAGT  
GATAGATAAGAAAGTAAGTCCCAGTAATTGGGATTATTCTCATTATTGTTCTGACAC  
ATCCTTTCACATTGACATGACAGATGGGCGTGGAAATAATGTCATTGCAAGAGAGGT  
AGATGCACTGATAACATCTCATGCGGTGATCCAGGGAAGGGAAACTAAGGGATGAAAG  
CAAGATACTGGCAGCAACAGAAGAATAACAGCATCCCCCTGAAATATGCACCCCC  
TATATATGTCCTACATATATATCCCCACACATATATCATCATTGTCACCTGTGAGACC  
TTTGTATCAATCTGGTTGTTGGGTTGGTTCTATTATAAAGATTTCTATTAGCC  
ACTATTGAATCTGAGATTGTATTCAAGAAAGGAAACGTCCTCCATCATCTCTGTGATGTC  
ACCCCAGAGAGAAGATTAGAAGATTAGCTGTCAGCTGACACCAGCCCAGCGGAGCG  
AAAATAATTGACATTCAAAGAGAATGAGGGAGAAATTCC

# 187

CTGCAGGGCTCTGTCATTGGAATGAGATTGGTTGCACTGTGGAGCTCTGTAGGAGGATG  
AGCTGTTTCTTAGTCAGGACCTACATTGAACCTGGGTGCTTCACAGGGTAGACCAGGG  
AATGAGGACCCCTCCTCAGTCCTCTTGATCATCTTCCCTCATTGCTTTGTGTCACAGC  
AGATGCTCCCTCTGATGTGCTGTTCTCCCTGTTCATACAGGACCATACACATGCTT  
CCAAGCTGGTAGTGCAGCACTGTGCTTCATTCACTGCATGATAGCCACAGAAAGAA  
ATATGACTTACAGGATATTGCTGGCAGCACGCATGCATTGCTATGTGCTTCAGITC  
CACAGATCTGATGTTCAAGGGTATCAGTTGAAAATTCTGCTTGGAGTTACATTTGCTCCAGAAC  
TATTTGCAACAGATAAAACATTCCCACACCGTGTGCTGCACCTCTGGAAAACGATGGTCTGAAT  
ATTAGAGACCAAATTCCCACACCGTGTGCTGCACCTCTGGAAAACGATGGTCTGAAT  
GGGGTTGTTCAAGATTACACCTGTGGAGCTAAGCGGAGCATTGCATGGAGGAAGGCCA  
AAACATATGGCAGTCCAATGTTGAGAGCAAATTAAATGCTTCCACCATTGATCT  
ACCGCGTTGTTAGGTTTATGGG

# 188

GGGTGGTCAACTACACTTAGGCTCACGTATCTCAGAAAAGTTCAAAACATTCTTCAGTA  
AAACTCATGAATGTTATCCAGCACTTCAGGATCCTCAGAGGCTTGATCCAAAACACGCT  
TAACTGCCTGTATTAAGGATAACAATTAAACCTTCATCTAAACACTTCATCCAAA  
ATTCTAAATCAGAGTTCAAGCCTTCATAAGTCACCAAGCTGGAAACCCATGAAG  
GATCAACAGATATGGTGTATTGTTGCCAAACAAACAAAGAATTAAATCTGGAACAATG  
ATCCTAAAAAGATGAAGAGGACACTGTTACTAAACAGTACACATATACGAAATAACTG  
TCAGGTGAAAGTAAGAACAAAGCAGGTTTCAGAAGACTGTATTAAATCAAATATG  
ATCAAATGTTACAGCAGCTTATTGGCATCTATAAATGGAACCACAGCTTTTTTTT

# 190

CTGCAGCAATTGCATCTATGGAAAGTATTAAATATATTAGCTATCTTGAATGTGAT  
CTGGCAGCTGGAAGCAAGGAGAGTCAGCACCATGTGAACAGTTAGCTTCCAGAGATAC  
CTAGCAACCATGTGAAGTTGAGCAATTCTGTTAGCCTAATTGCTAGTTACACCATT  
GCAGTGCTGTAGTTAGGCTGCTGTGATTCAAGTGTAGGAGACTTAACAAACGTT  
CTGTGCAACTACACACTGCTTCTTACTGTAGTAGGAGGGCGTAAGTAGTATGGTT  
AAGGTGTAAGGACTGGGGAGAGCTGGAGGAGAGTAATTCAAGCCTTGGGTTGGTGGGG  
AAGAGACCATAGCAGTTACTTGCCTCATGCTGGCTAATTCTTCTTCCAAACCTG  
CAAATTAGGAAGAAGGATTTGTACGGATGCAGGAGAAATCTCCGGATGCACTCTG  
CTTGAGAAGAATAGCAGCTTCATAAGTGAGATTAACTACATAATGCTGTTCTACTT  
TCCCATATACTTAAGTGTGTTAACAGATGCCAAGTGAGAATTGTTCAAGGCTGATTGG  
GTCTAGTAAGATTTCACAAGGGAAGG

# 191

GAGAGAGTCTGTATAAGATGATATTCAAGCTAAACTGACAACCTCCTATAGCAGTACaC  
CCTAACAGAAGATGAGCTATTAAACAGTGGAGAAGTGGAAAGAAAGATGCATGCTG  
GAAATTGTTAATAAAAGGATTCTGACATTGGAACATCATAGATTCTGCAAACCTGG  
GACAAGTTATTGACATCTTGTGATGAAAATATAGGCAGCTGAATCTCATGGATGC  
TAACCTCAAGATTAAATTGTTGCTCCACTGTGATTTACAGCTGAAATCTAGTGA  
TCTGTCAGCTGATCTGCTATTCTATTCAAAAGCCAGATGTAGCAGGAACAAgAGAC  
AAAAAAGCCGAAATAAATTCTCAGTGAAAGACACAGGGAAACAAAGGGAAATAGTG  
TTGTTCTCATGATATTAAAGACATCTTATTCAACATGATTCAATTAAACAAACTG  
GCAGGGCTTTGgAAACTAAAATGAGATATGAAAGAGGTACACTCACTCACTCAC  
ACCTACACCCACATACACACAGGAgGAAATCATTGACAGTAGCTGCAG

# 192

GTAAATGGACCGAAGTAAAAGAAAGGACAAAAAGACATTGGTAGTGTAGAAC  
CCATACAGCAGTACAAATGTTACATAATTGAGCACCACAAATTACTGGACAAGCTCAA  
TACGCATAAAATAGAAAGTTGCTGCTTAAGCAGCAGTGTGACCCCTTCCAAAGGCAG  
TAATAGAACCAAAAAATCTAAATTCTTAATGAGAATGTAATGCACTTCTTAAATAA  
ATAGAAAATAAATTAAAATCCACATAGAAAGAACAAAGTCCCCCTGAACCTAGGATAA  
ATTCTTGATGTCACCAAGGTTAAATGTAATAATTACATACAGATCGGTTATGAGT  
TAATAATAAAGATGCACAGCCTGCAG

# 193

ctgcaggtGtGAAAAAAtACtagtCTGtCcTTtTCaGGATtCtCAAAtCTtATTCTGTGAAttGctTTTCAg  
ATGCTtCaATAATTtTAAtCAAcCACTGttATGtTAgCtTTGtACTGtTGTTGTTACTTAgaTTTC  
ACATGAGaTTGGAgaGAaagAAAtgATTgTCCTGGgtAcaGGtaAAgAcTAgtGGAAGGAAGTAC  
NATGAGTGGtAAGTAAAGCtgaAGATGGCgagTcTTgTTGTATGGATCTGcgAAGAGATGAGG  
AaCAgTCCAAtGGGGtCTGGagCTCAGagtGgaGGAAGTTATCATAACTAATGGAGTCATTA  
GTGTCTAAAGCCATCCAGTAGGTTCTGTATCTGCTGGgACACACGAC

# 195

AGGAGCATATTACACATCCAGAACAGAGCTGAAGGAATAAAGCCATTCCATTGAATGGC  
TAAATTGCCATACAGTAAATAAACTAATAATTCAAAAATACGACACAGAGTGGAGA  
GAAAAAAACCCCTCCCCAGTCCCTGACTTCTCTCACTGGCCTTTTTTTTT

# 196

CACTGCTGGCAGCTATCACACTGCTTTCAGTATCACTTGCACACTATGTGTGCAAAAAA  
AACCCACCCCTACAAAGGAGATTCCAGCTGTGCTAAAGGTTGATACAACACTACTCGATG  
CTTCAAAGATAAGTAATGGAGGCTATTCTAACACAGATCTTAAATATGCCTCACACC  
TACTTCTCCTCACTCTGAACAGCTGTATATAACAAATGCCAGTGAGACTGCACAGACAC  
ATTTCCCTGACCATAACTCATTAGCACCAGAGACGCACTCCTTGAAAAAAATT  
TAGAGAAACACATTCCCCTCAACACTCGTTACATTCTATGGTACTAATCATTATTCA  
GCTAAATATAGTTACAATTTCAGTATCCCCTACCCACTACTCCCTTCGATG  
CTTGAACAGCAGTGAGATCTCCCTGGAGCTAGGATTAGTAAATGAGAAACAGAG  
CAGCTACCCCCCCCC

# 197

AGGTTCCAACCCCTGCCACAGGCAAGGCCACCAACCTCCATATCTGGTACTGGACCAGG  
CTGCCCAGGGCCCCATCCAACCTGGCCTGAACACCTCCAGGGATGGGCATCCACAGTC  
TCCCTGGCAGCTTCCAGGACCTCACCACTCTTCTGTGAAGAACTTCCACTGACATC  
CAATCTAAACCTCCCTCCTGAGCTTACCCCTCTGTTTTATTATGTGGAGGATCTTTTA  
GTGGTTGGGATAGGACCTCTGTTCTCCTTCTGTACACTGACAGAGGTTGTCA  
CACTATAGACAGGAATGCTATGCATACAAACCTCTGCTTGCACAGGTGCATAGCCAGAT  
GAGTGTACACAAATTCTTACCTTGAGCCCCAGGACCATTCAGGACTGCCAAAAGCAAC  
CACTGCTGTGCCAACACAGGAGCTGCTGGAAAATGCCATGCTCTGTGAGGAAAAAA  
ATCCGCCTGGGCTTGGAGCCATCAGCTGCCTAGAGCAGAGAAAACCAAGCTCCTCCC  
TGGGTTAGAGCAGCAGTACCTCCGAGCAGGAGAGGATGAAGCATGCCATGCACTGACA  
AAGCAGATTCTGCCTGACTGAGTAGCTGATGGTGCCTGCAAGGACCTGGCCAGAAGTG  
CACCCAAGGGTCTTGCATGAGTCCTGCCACCACTGCCATCCCTGAGAGGAGATGCA  
ACAGCACCTGCTGAGTGTGAGCTGGCAGCTAAAGGAGGTGGAGCAGCTGCTGTGCTT  
GCTTGTCTGCCAGACAGGAGCAGGAGCAGAGGGTAGGCTATGTATCAGATGGACTCCA  
CATCACCATAGTTCTGGAGGACTGAAAGACTGTGCTAGCACACCCCCACCCCTATGCA  
CAGGTGCAGATGATGTGCTGGATACAACACTGGAGCTGCAG

# 198

CtGCAGAGTTCTGACTCAAAAGCCAGGGACCTAGGTAAGAATCAATAACAGCTGACACA  
CAGCAACGTGGCTATGACTGAAAATGAAAGTGTTCAGTACTTTGTCCTGCTGTATCTT  
CAAAGGTAAATTAAATTAGATGTAGGTGCCAGGAAGAGAAGTAAACTGACATCAA  
CTCGTTACATCAGATGCTAAATAGCATTGTGAAGTAACAAGCTTCTCTCCACTAG  
CCAGACTGTTCTAGATTACACATGCTGCTTCTGGTATAGTCTACCATGAAAGCTGATCT  
ATTATTGGCCTAGATTGCTTACATTCTAAATTGTGCCATGGTATCCCTGAAACTTT  
GTTTATGAGCTGGAAGAACCTGTGATTAATTGGCCCTTTATTAGTGCCTAGTCAG  
AAAAATATTCTGCTTAATGAAATAGAAGTACTTCCCTGAAATGCCAGTTGTCTAATT  
ATACTCCCATCTACAAGGATACAAGGCTACTCATCACACAAAATCTCACCCAAACCAC  
TGGAAAGTTAAAGGTCTACAAATGGACAACGGAGTCAGGGAGGAAACACATGTGCTCA  
ATTAACCTCAACAGATTAAGAATAGTTGAGTACATCAGCCTAAAGGTGTTCTCTGT  
ACTGGATTTAACGATACAAAGCACTGCTGCCATTACATTATGTTTTGTGTTAAGTT  
TTGGGTAGATAAGGACAGTTAGCAGAGAGCTAACAGACGCTTACAGGGAAACTCAGGC  
ACAAGAGGTTTGAGTTGCCACAAGTCCAAGACCAAGTCAGTCACTGATCTGGCATTCT  
GATCAACATTGTTACCCCTTTAAAAGGCAAATTAGAATCCTGAGTTATGCAATAACTTT  
AATTAAACACCGTCAAGAATATTATTCAATAGCGATACTAATAGCTAATAGTTAATAGC  
TAATAACTAATAGTTAACCTACCTAATAGTAATTAGCA

# 199

CTGCAGAAACATTGGTTCTTTATTTATAAAAATCATTCTATGTTGGCATACTT  
ATTAATATGAAATAGAAATATTCATTCAACTTCCAAAATGGATAGGAAGTACAGCT  
CTGCTAAAAAAATGTCAGAATGTAAGAAAAATTATTTGGACTTACACTGCCTGTA  
AATAGTTAGGTCTTACTGGTATTGTTATCTCACACAGCAGACATAAAAGGCTAACAG  
CCTACATAACAGTGAGACTGCCAGCACAATGAGGACAGCTCATTAATGGGAGCTGAGT  
ACTCTACTGTGCTGAGATATGACATGTCAGAAGAATGTCAGAATAACTGAGCAGGTGAC  
AATTAAGTCTCAGTTGCTACTGCTCTCCTGGTTATTGGAGGGTTGGAGTCAGAAGA  
CAGCTGTTACACTCTTCCAGGGCTTGCTTAGCT

# 200

TCTTAGTCTGTTGGTAGTGAAGTAAAAGGTCACTTGATAGAGAACATGGGATGGCTT  
AGTCGTTGTTGAGGACATGTAACCTTGCTCTGGCTTAGTCCAAGTCAAACCATATAAG  
TACATATATAAGTGCATCAGATCTATGAGTTTAGTGTCAAGTGAAGTTGAGTGTGCT  
GTGACAATCCCAGTATAAGTGAACCTATCGTAACAGCTAGCATGGCGCAAGTTA  
TCTGCTCTAGCCAGCTAGGTACACATGAAAACATTACCCACACACTGACTCACACACTGC  
AACCTCTTACTCATGCAAGCATACTTGATGACAGTTGTTACTAATAACACAAGCTCAC  
ATTCAGAGATCAATTATATTAGCCATGTGCGTACCCAAAACCTCCTGAGCAATGTGA  
ACTGAAGATATCCTTTATGGAGGGCTACTGATAGGTTAAAGACCTTATATAAAAACC  
CTGAGACGTGAAGGTAATCTGAGTCTCAGAGATGTGAACCTGGATTACCAACTCTGGTG  
AGATTTTTAAACTATTTTCTTTGATACCCCTCAGCAATGCAAAATCATTGATTGTC  
GTCTGTTATCTCAGGATAAAAGATAAGTTGTAATCCTAGGCTACAGGAGGAGAGTTG  
GAAGACGCAGTCTTGTAAACTTGATATTtGACTGGTGGCAATGCTCTACCTAAACCT  
CAATACCAAATAAGACTGGTGTGTTAGACCACTGCTGTTGAATTGCAACTG  
TGCTCTCAATAATCACCTTGAAGTGCAGCATTCCACTGCACTCAAGAACGAGACCA  
AGAAAAGGATTGTAAGAGCTGTACAGCTCTGGCTACCTAGTACATtCAaGCAGA

# 201

GATGCCTACATGATGTGCCAGAGCCCTGCCCCAGGATGGGTTATCCACCACTCTG  
GACAACCTGTGCCAGTGCATCACCAACCTTATTGTAAAAACACTTTTATATCCAATCTA  
AATCTCTCCTCGTATAGTTGAAACCGTTGCCCTTGTCCATCAACAATTAGATTGGTCT  
AATCTCAATTGCCCTTCATCACATACTCTGGGTGCATCTGATCTCGAGCCTGGAGCTA  
AAAAACCACTGAACTGATTTCTGTGAATAAAAGTTCAAGGTTGCATTGCCAAATGG  
AGCTTTCTGGTCTCTCAAACCTCCCTGGAATAGAAATTCTTCAGCACAGTCT  
CTCGAGACCACCTTTCTAGAATGCACAGGCTGCTCTGGATCTGAGGAAAGGCCA  
ATTTAAGATCAAGCAGCAGGTACGTTGCAAGTGAATATTCAATTCTTC  
CTCTCGTTAGCGTAATGCCCTGATCAGTGCCTCAAGCACTACTGCAG