

Ćwiczenie C03 Mobilom człowieka

Metody analizy transpozonów Transpozony w liniach rozwojowych Aktywność transpozonów u człowieka

Kornelia Polok

1. Metody analizy transpozonów

➔ Transpozony występują powszechnie w genomach Eukariota, w tym w genomie człowieka. Transpozony mogą być aktywowane pod wpływem stresu, wpływają na adaptację do zmieniających się warunków środowiska i stanowią siłę napędową ewolucji. Transpozony są także wygodnym narzędziem w badaniach podstawowych oraz inżynierii genetycznej. Z drugiej strony, duży udział transpozonów w genomach utrudnia sekwencjonowanie i składanie sekwencji. Tradycyjne metody identyfikacji transpozonów obejmują analizy genetyczne i molekularne. Dostępność sekwencji genomowych w bazach danych umożliwia także identyfikację w oparciu o analizy bioinformatyczne.

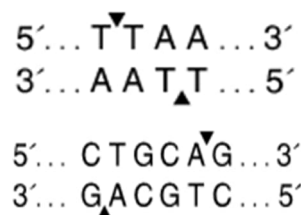
1.1. Identyfikacja miejsc insercji transpozonów

Transpozony należą do rodzin, które charakteryzują się określonymi motywami strukturalnymi. Motywy te można wykorzystać do przeszukiwania sekwencji genomowych oraz do projektowania starterów. Przykładowo, metoda SSAP (ang. Sequence-Specific Amplification Polymorphism) wykorzystuje startery komplementarne do sekwencji transpozonu i pobliskiego miejsca restrykcyjnego. Amplifikacja w obrębie insercji przejawia się w postaci prążka na żelu. Dodatkowo, wykorzystanie enzymów restrykcyjnych wrażliwych na metylację DNA pozwala śledzić zmiany epigenetyczne związane z dynamiką transpozonów.

1.1.1. Podstawy techniki SSAP

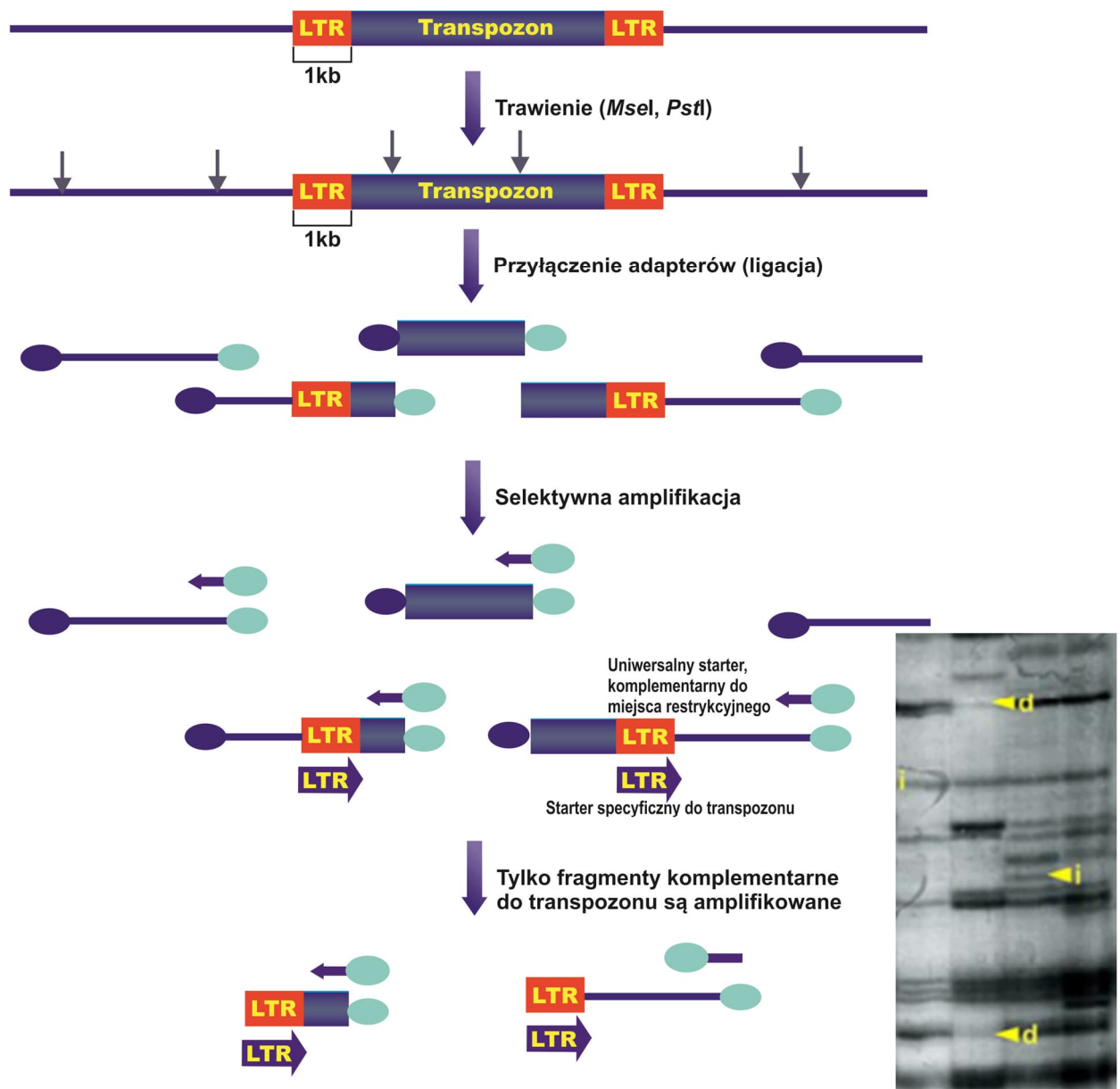
• Genomowy DNA jest trawiony enzymami restrykcyjnymi:

- ▶ *Mse*I, który rozpoznaje sekwencję: TTAA;
- ▶ *Pst*I, który rozpoznaje sekwencję CTGCAG.



Rys. 1.1.1a. Sekwencje rozpoznawane przez *Mse*I (górną) i *Pst*I (dół).

- W wyniku cięcia enzymami restrykcyjnymi otrzymujemy fragmenty, które zawierają miejsce restrykcyjne oraz transpozon.
- Otrzymane fragmenty mają lepkie końce. Przyłączamy do nich adaptory o ustalonej sekwencji nukleotydowej oraz komplementarne do lepkich końców powstałych w wyniku cięcia enzymami restrykcyjnymi.
- Dodajemy startery:
 - ▶ Jeden starter jest komplementarny do adaptera. Dodatkowo wprowadza się 1-3 losowe zasady, aby uwzględnić polimorfizm w miejscu insercji.
 - ▶ Drugi starter jest komplementarny do sekwencji transpozonu.
- Dzięki doborowi dwóch starterów amplifikowany będzie fragment zawierający transpozon oraz sekwencję przylegającą do miejsca insercji.



Rys.1.1.1b. Schemat analizy miejsc insercji transpozonów.

- Obecność transpozonu można identyfikować za pomocą prążków na żelu lub w drodze sekwencjonowania. Miejsce insercji jest uznawane za nowe, jeżeli w danej pozycji kontrola nie miała prążka. Brak prążka istniejącego w kontroli oznacza delecję.



1.1.2. Identyfikacja transpozonu *SIRE* w genomie człowieka

Niektóre transpozony są ograniczone do specyficznych linii rozwojowych, podczas gdy inne występują w wielu liniach ewolucyjnych. Przykładowo, wykryty u jęczmienia retrotransponon *BARE-1* z grupy *Ty1/copia* występuje u pijawek (dane własne), koni i owiec (Elkina i inni 2015). Z kolei u człowieka wykazano obecność retrotranspozonu *SIRE1* pochodzącego z soi (Guner i inni 2022).

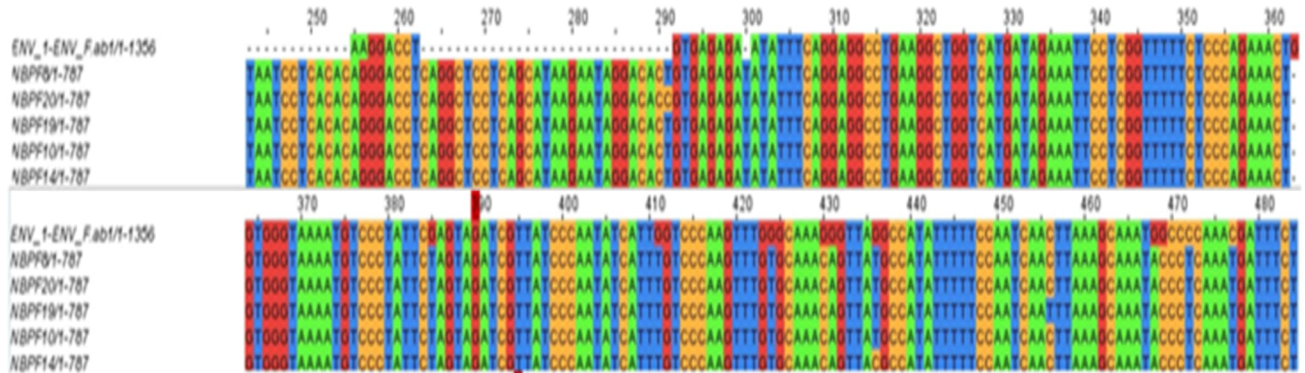
Retrotransponon *SIRE1* należy do transpozonów *LTR*, grupy *Ty1/copia*. Każda kopia ma wielkość około 11 kbp. Cechą charakterystyczną jest obecność dodatkowego ORF, kodującego białko podobne do białek otoczki retrowirusów.

- A. W tabeli 1.1.2 przedstawiono wyniki analizy genomu ludzkiego za pomocą starterów komplementarnych do sekwencji transpozonu *SIRE1*. Na podstawie przedstawionej tabeli proszę podać średnią liczbę insercji i delecji *SIRE1* w genomie ludzkim z uwzględnieniem całej populacji oraz z uwzględnieniem podziału na płcie. Czy pomiędzy płciami istnieje różnica związana z transpozonom *SIRE1*? Czy istnieje różnica między liczbą insercji i delecji? W obliczeniach proszę wykorzystać stronę: <https://www.statskingdom.com/two-way-anova-calculator.html> lub program Statistica na komputerze.

| Tabela 1.1.2. Miejsca insercji transpozonu <i>SIRE1</i> w genomie ludzkim | | | | | |
|---|----------|---------|-----------|----------|---------|
| Kobiety | Insercje | Delecje | Mężczyźni | Insercje | Delecje |
| 1 | 12 | 4 | 1 | 16 | 0 |
| 2 | 12 | 4 | 2 | 16 | 0 |
| 3 | 12 | 4 | 3 | 16 | 0 |
| 4 | 10 | 6 | 4 | 14 | 2 |
| 5 | 12 | 4 | 5 | 14 | 2 |
| 6 | 14 | 2 | 6 | 12 | 4 |
| 7 | 15 | 1 | 7 | 12 | 5 |
| 8 | 16 | 0 | 8 | 12 | 4 |
| 9 | 16 | 0 | 9 | 12 | 3 |
| 10 | 14 | 2 | 10 | 10 | 4 |
| 11 | 14 | 2 | 11 | 12 | 4 |
| 12 | 16 | 0 | 12 | 11 | 4 |

- B. Sekwencjonowanie wykazało, że sekwencje *SIRE1* są zlokalizowane w chromosomach 1, 4, 9, 20 i 22. W obrębie chromosomu 1 sekwencje *SIRE1* zidentyfikowano w obrębie genu *NBPF3*. Ekspansja rodziny genów *NBPF* miała miejsce podczas ewolucji naczelnych i związana była z rozwojem mózgu. Liczba kopii genów *NBPF* jest zmienna w populacji ludzkiej. Geny te charakteryzują się obecnością tandemowych powtórzeń, co może być związane z ich retrotranspozonomowym pochodzeniem.

Rysunek 1.1.2 przedstawia uliniowanie sekwencji *SIRE1* o długości 242 bp z genami rodziny *NBPF*. Proszę zidentyfikować wszystkie mutacje, które zaszyły w *SIRE1* w trakcie ewolucji naczelnych. Proszę uwzględnić tylko te mutacje, które występują we wszystkich genach *NBPF*. Na podstawie zidentyfikowanych mutacji proszę podać procent zasad, które uległy zmianie w *NBPF*.



Rys.1.1.2. Uliniowanie sekwencji transpozonu *SIRE1* oraz genów *NBPF* u człowieka.

1.2. Identyfikacja sekwencji transpozonowych

➔ 1.2.1. Odwrócone powtórzenia terminalne (TIR)

Odwrócone powtórzenia terminalne (TIR) to sekwencje zlokalizowane na końcach większości transpozonów DNA u Eukariota. Występują one także w elementach insercyjnych Prokariota. Elementy insercyjne są najprostszymi transpozonami, które zawierają tylko geny niezbędne do przemieszczania. Jeżeli dwa elementy insercyjne są blisko siebie, to tworzą one transpozon złożony. Transpozony DNA zawierające TIR rozpoznaje się na podstawie obecności odwróconych powtórzeń na końcach elementu (Rys. 1.2.1).

5'CTGACTCTT ————— AAGAGTCAG3'

3'GACTGAGAA ————— TTCTCAGTC5'

Rys.1.2.1. Odwrócone powtórzenia terminalne (TIR) na końcach elementów insercyjnych.

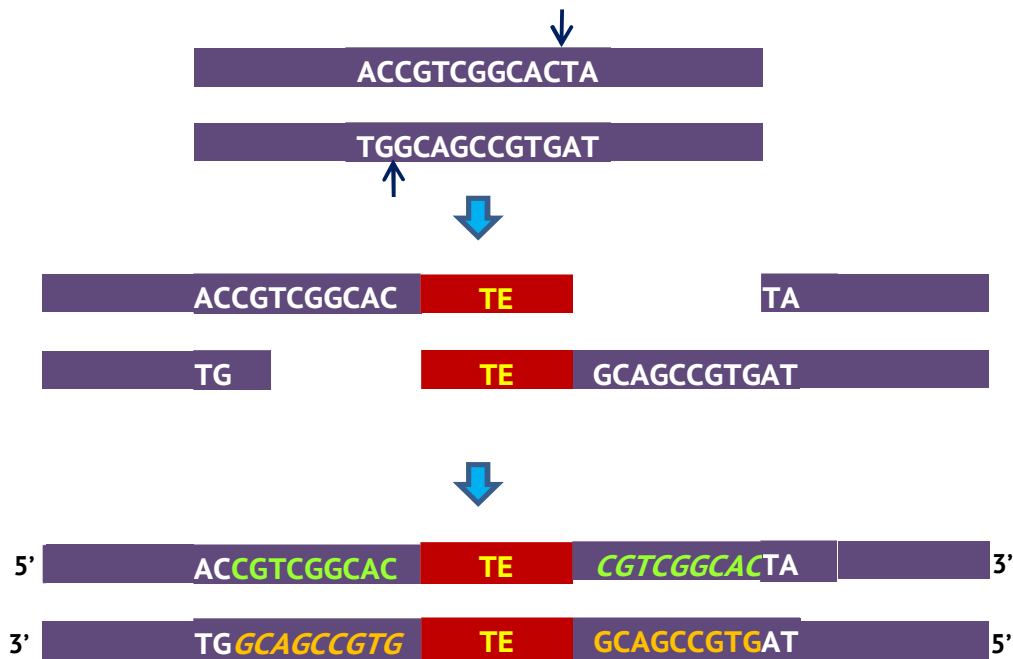
1.2.2. Duplikacje w miejscu insercji transpozonu

Gdy transpozon wbudowuje się do genomu to w miejscu insercji dochodzi do duplikacji sekwencji. Sekwencje te mają długość 2-13 bp. Flankują one transpozon. Prawdopodobnie są efektem asymetrycznego przecinania akceptorowego DNA podczas wbudowywania (Rys. 1.2.2).



1.2.3. Która para sekwencji DNA może zostać zakwalifikowana jako TIR przy założeniu, że przedstawiają one sekwencję na tej samej nici DNA? Proszę wyjaśnić.

- A. 5'GAATCCGCA3' i 5'ACGCCTAAG3'
- B. 5' GAATCCGCA3' i 5'CTTAGGCGT3'
- C. 5' GAATCCGCA3' i 5'GAATCCGCA3'
- D. 5' GAATCCGCA3' i 5'TGCGGATTC3'



Rys.1.2.2. Powstawanie duplikacji w miejscu insercji transpozonu. Kursywą zaznaczono dosyntetyzowaną sekwencję. Kolor oznacza powtórzenie.



1.2.4. Która para sekwencji wskazuje na duplikację w miejscu insercji przy założeniu, że przedstawiają one sekwencję na tej samej nici DNA? Proszę wyjaśnić.)

- A. 5'AATTCGCGT3' i 5'AATTCGCGT3'
- B. 5'AATTCGCGT3' i 5'TGCGCTTAA3'
- C. 5'AATTCGCGT3' i 5'TTAAGCCGCA'
- D. 5'AATTCGCGT3' i 5'ACGCGAATT3'

1.3. Identyfikacja transpozonów metodami bioinformatycznymi



Metody bioinformatyczne znajdują zastosowanie gdy nie posiadamy wiedzy o sekwencji transpozonu. Wówczas poszukuje się charakterystycznych powtórzeń w parach podobnych sekwencji w różnych lokalizacjach, a następnie przeprowadza się grupowanie, aby scharakteryzować rodzinę powtórzeń. Metody te związane są z trudnościami wynikającymi z:

- rozproszonego charakteru sekwencji transpozonowych, w tym lokalizacji sekwencji transpozonowych w innych powtórzeniach;
- fragmentacji sekwencji transpozonowych;
- trudności z rozróżnieniem blisko spokrewnionych rodzin, które mają wspólną znaczną część sekwencji.

Analiza transpozonów *de novo* obejmuje kilka etapów.

- Poszukiwanie regionów powtarzalnych w genomie (np. RepeatFinder).
- Grupowanie znalezionych sekwencji i oddzielenie sekwencji, które nie pochodzą od transpozonów.
- Analiza homologii pomiędzy daną sekwencją a znanymi sekwencjami transpozonów. Analizę najczęściej wykonuje się na poziomie białek. Metody te wykazują odchylenie w kierunku już znanych rodzin transpozonowych. Nie nadają się one do poszukiwania niektórych transpozonów, np. *SINE* u człowieka.

- Metody porównujące cały genom polegają na poszukiwaniu regionów zawierających insercje, np. regiony ortologiczne są przerwane insercjami o długości >200 bp. Metoda użyteczna przy wykrywaniu nowych rodzin transpozonych.

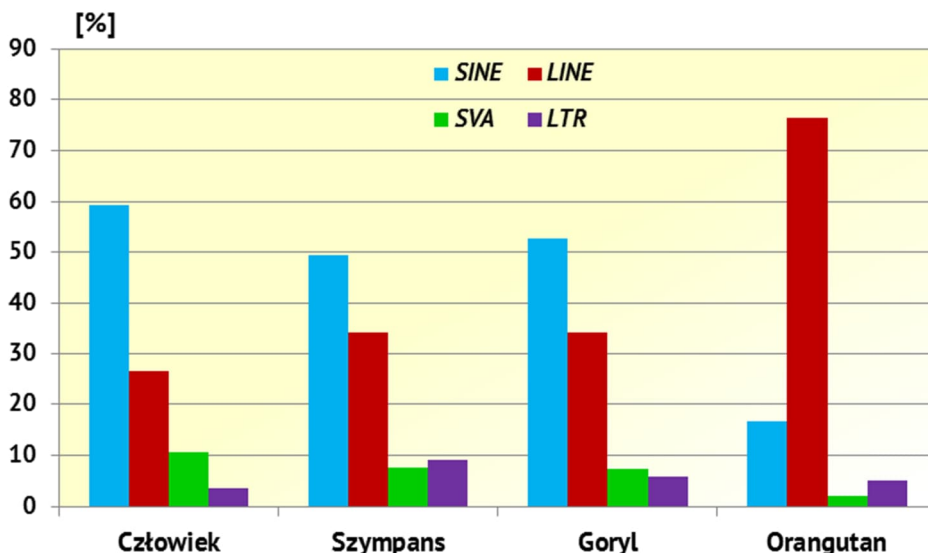
2. Transpozony w liniach rozwojowych

2.1. Specyfika mobilomu Hominidae

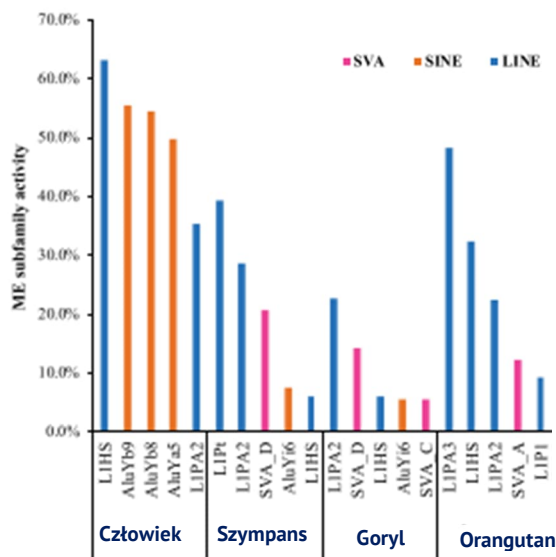
➔ Porównanie mobilomu człowieka, szympansa, goryla i orangutana wskazuje, że każdy z tych gatunków ma specyficzny skład transpozonów. Analiza bioinformatyczna zsekwencjonowanych genomów zidentyfikowała 2-3 razy mniej

transpozonów u orangutana (6 437) niż człowieka (12 081), goryla (15 365) i szympansa (17 379). Różnice występują także w udziale poszczególnych klas retrotranspozonów w ogólnej liczbie transpozonów (Rys. 2.1a). U człowieka, podobnie jak u szympansa i goryla najliczniejszą klasą są nieautonomiczne elementy nie-LTR, *SINE*. Drugą najliczniejszą grupę stanowią transpozony *LINE*. Cechą charakterystyczną człowieka jest nieco wyższy udział transpozonów *SVA* niż u pozostałych Hominidae. Natomiast genom orangutana charakteryzuje się całkowicie odmiennymi proporcjami retrotranspozonów.

Różnice dotyczą także aktywności transpozonów. U człowieka wysoką aktywność wykazują retrotranspozony *SINE* (głównie *ALU*) i *LINE*. U pozostałych Hominidae aktywności *LINE* i *SINE* są niższe (Rys. 2.1b). Różnice w składzie i aktywności transpozonów niewątpliwie wpływają na różnice funkcjonalne i regulacyjne między genomami Hominidae.



Rys.2.1a. Porównanie udziału różnych klas retrotranspozonów w ogólnej liczbie transpozonów w genomach Hominidae (na podstawie Tang i Liang 2019).



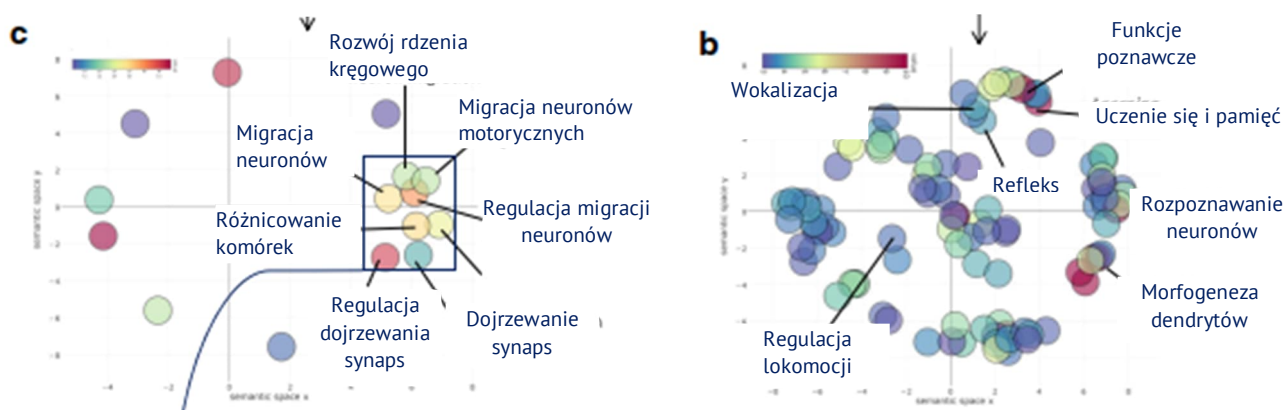
Rys.2.1b. Porównanie aktywności różnych klas retrotranspozonów w genomach Hominidae (Tang i Liang 2019).

2.2. Retrotranspozony w ewolucji współczesnego człowieka

Retrotranspozony typu nie-LTR są głównym czynnikiem, który doprowadził do specjacji *Homo sapiens*, dostarczając zmienności niezbędnej do ewolucji, w tym do powstania złożoności neurologicznej. Częstość insercji transpozonów nie-LTR w linii ludzkiej jest 2,5 raza wyższa niż w linii szympanów. W genomach *Homo neanderthalensis* oraz *H. denisova* i współczesnego *H. sapiens* zidentyfikowano łącznie 102 insercje *Alu*, *LINE-1* i *SVA*, które nie występują u szympanów. W przypadku *H. sapiens* insercje te wykazują polimorfizm, przy czym populacje afrykańskie na ogół nie mają tych insercji *Alu*, *LINE-1* i *SVA*. Obecność insercji w populacjach nieafrykańskich potwierdza znany fakt przepływu genów między populacjami *Homo* na terenie Eurazji.

Ponadto w genomie współczesnych populacji *H. sapiens* występuje 321 specyficznych insercji, które nie występowały w genomach *H. neanderthalensis* oraz *H. denisova*. Większość specyficznych insercji dotyczy elementów *Alu* (295 = 92%), natomiast znacznie mniej jest specyficznych insercji *LINE-1* (25 = 7,7%). Insercje te wykazują polimorfizm. Około 7,5% insercji jest powszechna we wszystkich populacjach ludzkich (częstość insercji >85%). Około 2,5% insercji jest polimorficzna w populacjach afrykańskich (obecność insercji <65%) i bardzo częsta w populacjach nieafrykańskich (obecność insercji >85%), co sugeruje, że utrwalenie się tych insercji było związane z migracją z Afryki. Rozkład insercji w populacjach ludzkich odzwierciedla historię migracji z generalnie większym zróżnicowaniem populacji afrykańskich.

Insercje specyficzne dla współczesnych populacji *H. sapiens* występują głównie w genach o dużym zróżnicowaniu transkryptów oraz wykazujących zróżnicowaną ekspresję w tkankach. Najczęściej insercje związane są z genami, których ekspresja ma miejsce w mózgu, w szczególności w ciele migdałowatym, hipokampie i niezróżnicowanych neuronach (52% insercji). Dane te wskazują, że wzrost zróżnicowania transkryptów na skutek insercji był pozytywnie skorelowany z funkcjami kognytywnymi charakterystycznymi dla człowieka.



Rys.2.2. Funkcje genów, w których występują insercje transpozonów specyficzne dla *H. sapiens* (Guichard i inni 2018).

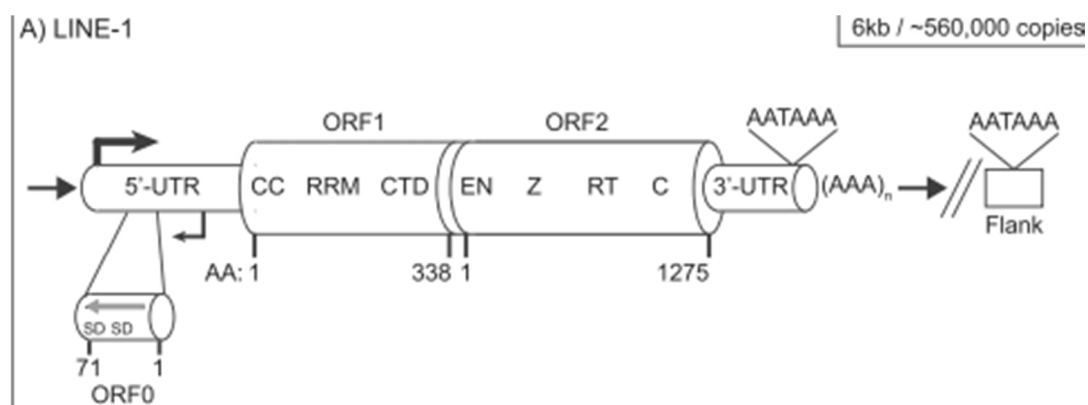
Przykładem insercji specyficznych dla współczesnego człowieka jest obecność *AluYg6* w chromosomie 1 (chr1q25.3) w obrębie genu *EDEM3* kodującego enzym uczestniczący w degradacji białek w reticulum endoplazmatycznym oraz *AluYb9* w chromosomie 10 (chr10q25.3) w obrębie genu *SHTN1*, który koduje białko wiążące się z pirymidyną. Insercja w *EDEM3* ma miejsce w

obrębie intronu i prowadzi do jego egzonifikacji. Insercja w *SHTN1* występuje w intronie, prowadzi do skrócenia transkryptu na skutek utraty dwóch pierwszych egzonów. Gen *SHTN1* ulega silnej ekspresji w ludzkim mózgu i jest odpowiedzialny za wytwarzanie sygnałów niezbędnych do wzrostu i polaryzacji neuronów.

3. Aktywność transpozonów u człowieka

3.1. Transpozony aktywne w genomie współczesnego człowieka

➔ Niektóre transpozony są nadal aktywne u człowieka. Dotyczy to około 80-100 elementów *LINE-1* spośród 0,5 mln elementów obecnych w genomie. Każdy człowiek ma nieco inny zestaw aktywnych elementów *LINE-1*. Aktywny element ma 6 kbp długości, zawiera miejsca 5'UTR oraz 3'UTR, geny dla dwóch białek: ORF1p i ORF2p, które oddzielone są spacerem o długości 63 bp, a



Rys.3.1. Budowa aktywnego elementu *LINE1* u człowieka (Hanks i Kazazian 2016).

także łańcuch poliadenylowy (Rys. 3.1).

- *ORF1* koduje białko ORF1p o masie 40 kDa, które wykazuje aktywność białka RBP (białko wiążące się z RNA) oraz aktywność białka opiekuńczego. Białko tworzy serię trimerów z kwasami nukleotydowymi na skutek szybkiej poliploidyzacji koordynowanej przez superskręcone domeny (coiled-coiled). Fosforylacja białka ORF1p jest niezbędna do transpozycji.
- *ORF2* koduje białko o masie 150 kDa, które wykazuje aktywność endonukleazy i odwrotnej transkryptazy.
- *ORF0* koduje niedawno wykryty peptyd o długości 70 aminokwasów. Peptyd posiada dwa miejsca donorowe, dzięki czemu może tworzyć różne produkty podczas alternatywnego splicingu, np. białko może powstać na skutek fuzji alternatywnych form. Ekspresja *ORF0* w warunkach *in vitro* prowadzi do wzrostu transpozycji o 41%.
- **Ekspresja *LINE-1***
 - ▶ *LINE-1* jest transkrybowany z własnego promotora przez polimerazę RNA II. Promotor zlokalizowany jest w fragmencie 5'UTR o długości 900 bp. Ponadto w regionie 5'UTR znajduje się dodatkowy promotor w orientacji antysensownej.
 - ▶ W trakcie transkrypcji dodawana jest „czapeczka” CAP.
 - ▶ W inicjacji transkrypcji uczestniczą czynniki transkrypcyjne: YY1 (ying yang), TCF/LEF (wzmacniacz w limfocytach T), RUNX3.

- ▶ Koniec transkrypcji związany jest z zewnętrznym słabym sygnałem polyA, AATAAA, który zlokalizowany jest w obszarze 3'UTR o długości 200 bp.
- ▶ W wyniku transkrypcji *LINE-1*, powstaje RNA, który jest transportowany do cytoplazmy, gdzie dochodzi do translacji ORF1p i ORF2p i złożenia transpozonu.
- ▶ Złożony *LINE-1* składa się z trimerów ORF1p, ORF2p, *LINE-1*-RNA, a także białek nie należących do *LINE-1* oraz komórkowego RNA. Częsteczki te są transportowane do jądra, gdzie dochodzi do odwrotnej transkrypcji i integracji.
- ▶ Miejsce integracji jest determinowane przez właściwości ORF2p, białka które najczęściej wykrywa sekwencję konsensusową 5'TTTT/AA3'.
- ▶ Obok mobilizacji własnego RNA, *LINE-1* prowadzi do retrotranspozycji wielu typów RNA komórkowego, w tym elementów *Alu*, *SVA* oraz RNA kodowanych przez genom – U6 RNA oraz mRNA dla białek. Genom człowieka zawiera około 8 tys. pseudogenów, które są retrokopiami i obecnie podlegają ewolucji.

3.2. Choroby człowieka wywołane insercją transpozonów

Zidentyfikowano około 124 jednostki chorobowe na skutek insercji transpozonu w czasie współczesnym. Insercja do egzonu lub intronu, który podlega splicingowi może prowadzić do przesunięcia ramki odczytu i powstania wadliwego białka. Nowe insercje związane są z delecją kilku do milionów nukleotydów w miejscu wstawienia. Dodatkowo insercja transpozonu może spowodować zmianę ekspresji genu. Dane z sekwencjonowania genomów, zwłaszcza sekwencjonowania pojedynczych komórek wskazują, że insercje w somatycznych komórkach mogą być częste.

- Insercja elementu *SVA* do genu *FKTN* kodującego fukutynę prowadzi do alternatywnego splicingu. W efekcie powstałe białko jest zlokalizowane w retikulum endoplazmatycznym zamiast w aparacie Golgiego. Zmiana ta prowadzi do powstania dystrofii mięśniowej.
- Nerwiakowłokniakowatość typu I związana może być z insercją *SVA* do genu *NF1*. Insercja związana jest z delecją od 876 do 1 000 kbp. Elementy *SVA*, które uległy insercji pochodzą z chromosomu 6 lub 10.
- Zespół Rotoru związany z obecnością nadmiaru bilirubiny, która jednak nie gromadzi się w wątrobie. Choroba występuje u homozygot względem insercji pełnej długości transpozonu *LINE-1* do intronu 5 genu *SLCO1B3*. Gen ten koduje transporter anionowy specyficzny dla wątroby. Insercja prowadzi do nieprawidłowego splicingu i w efekcie utraty aktywności białka *SLCO1B3* w wątrobie. Mutacja ta występuje u około 6% populacji japońskiej. Zespół Rotoru jest syndromem dwugenowym, dlatego sama insercja nie jest wystarczająca dla jego wywołania. Występuje ona tylko u osób, które dodatkowo mają mutację nonsensowną w stanie homozygotycznym w genie *SLCO1B1*.
- Insercja pełnej długości elementu *LINE-1* do intronu 14 genu *RB1*, który koduje białko zatrzymujące cykl komórkowy. Niefunkcjonalne białko prowadzi do niekontrolowanego namnażania komórek i powstania nowotworu, siatkówczaka. Dlatego gen *RB1* określany jest mianem genu supresorowego, czyli zapobiegającego rozwojowi nowotworów. Insercja *LINE-1* prowadzi do powstania niefunkcjonalnego białka i w efekcie do rozwoju siatkówczaka.



3.2.1. Hemofilia typu A jako efekt insercji *Alu* do genu kodującego czynnik FVIII.

Jedną z przyczyn hemofilii typu A jest insercja transpozonu *Alu* do genu kodującego czynnik FVIII. Częstość insercji transpozonu *Alu* w gametach wynosi 5%. Częstość występowania hemofilii A u mężczyzn w Polsce wynosi 1/5000. Gen kodujący czynnik VIII ma długość 186 kbp. Czy obserwowaną częstość hemofilii A u mężczyzn w Polsce można wytłumaczyć tylko nowymi insercjami transpozonu *Alu*? Proszę uzasadnić przyjmując, że prawdopodobieństwo insercji jest identyczne dla całego genomu człowieka.

4. Literatura

- Elkina MA., Erkenov TA, Glazko VI. 2015 Mobile genetic elements as a tool for the analysis of genetic differentiation of varieties of cultivated plants and breeds of farm animals. *IJRSR* 6: 5893–5900.
- Guichard E, Peona V, Tagliacucchi GM, Abitante L, Jagoda E, Musella M, Ricci M, Rubio-Roldán A, Sarno S, Luiselli D I inni. 2018. *Mobile DNA* 9:28. Dostęp: <https://doi.org/10.1186/s13100-018-0133-4>
- Guner BC, Karlik E, Gozukirmizi N. 2022. Remnants of *SIRE1* retrotransposon in human genome. *Journal of Genetics* 102:10. Dostęp: <https://doi.org/10.1007/s12041-022-01398-3>
- Hancks DC, Kazazian HH, Jr. 2016. Roles for retrotransposon insertions in human disease. *Mobile DNA* 7:9 Dostęp: DOI 10.1186/s13100-016-0065-9
- Tang W, Liang P. 2019. Comparative genomics analysis reveals high levels of differential retrotransposition among primates from the Hominidae and the Cercopithecidae families. *GBE*. Dostęp: <https://academic.oup.com/gbe/article/11/11/3309/5607297>

Odpowiedzi

1. Metody analizy transpozonów

1.1. Identyfikacja miejsc insercji

1.1.2. Identyfikacja transpozonu *SIRE1* w genomie człowieka

A. W tabeli 1.2.2 przedstawiono wyniki analizy genomu ludzkiego za pomocą starterów komplementarnych do sekwencji transpozonu *SIRE1*. Na podstawie przedstawionej tabeli proszę podać średnią liczbę insercji i delecji *SIRE1* w genomie ludzkim z uwzględnieniem całej populacji oraz z uwzględnieniem podziału na płcie. Czy pomiędzy płciami istnieje różnica związana z transpozonom *SIRE1*? Czy istnieje różnica między liczbą insercji i delecji? W obliczeniach proszę wykorzystać stronę: <https://www.statskingdom.com/two-way-anova-calculator.html> lub program Statistica na komputerze.

| Tabela 1.1.2. Miejsca insercji transpozonu <i>SIRE1</i> w genomie ludzkim | | | | | |
|---|-------------------------|------------------------|-----------|-------------------------|------------------------|
| Kobiety | Insercje | Delecje | Mężczyźni | Insercje | Delecje |
| 1 | 12 | 4 | 1 | 16 | 0 |
| 2 | 12 | 4 | 2 | 16 | 0 |
| 3 | 12 | 4 | 3 | 16 | 0 |
| 4 | 10 | 6 | 4 | 14 | 2 |
| 5 | 12 | 4 | 5 | 14 | 2 |
| 6 | 14 | 2 | 6 | 12 | 4 |
| 7 | 15 | 1 | 7 | 12 | 5 |
| 8 | 16 | 0 | 8 | 12 | 4 |
| 9 | 16 | 0 | 9 | 12 | 3 |
| 10 | 14 | 2 | 10 | 10 | 4 |
| 11 | 14 | 2 | 11 | 12 | 4 |
| 12 | 16 | 0 | 12 | 11 | 4 |
| Total K+M | 13,6^a | 2,4^b | | 13,1^a | 2,7^b |
| | 13,3^a | 2,5^b | | | |

- Należy przeprowadzić dwuczynnikową analizę wariancji dla obu płci. Wprawdzie dane nie spełniają warunku normalności zgodnie z testem Shapiro-Wilk, jednakże liczba ich jest większa niż 30 i wartości rozmieszczone są symetrycznie względem średniej, co pozwala na zastosowanie analizy wariancji. Alternatywnie można przeprowadzić testy nieparametryczne.
- Wchodzimy na stronę <https://www.statskingdom.com/two-way-anova-calculator.html> i w tabeli wpisujemy jako zmienną niezależną płeć, jako zmienne zależne liczby insercji i delecji. Odpowiednio wprowadzamy dane.

| Var A \ Var B | B-Insercje | B-Delecje |
|---------------|----------------|-------------|
| A-Kobiety | 12 12 12 | 4 4 4 |
| A-Mężczyźni | 10 12 11 | 4 4 4 |

Rys.1.1.2a. Zrzut ekranu ilustrujący wprowadzenie danych do tabeli w dwuczynnikowej analizie wariancji.

- Korzystamy z przycisku „calculate”. Otrzymujemy tabelę z wynikami ANOVA i ich objaśnieniem.

| Source | DF | Sum of Square (SS) | Mean Square (MS) | F Statistic (df ₁ ,df ₂) | P-value |
|------------------------|-----------|--------------------|------------------|---|---------|
| Factor A - rows (A) | 1 | 0.2 | 0.2 | 0.05 (1,44) | 0.8 |
| Factor B - columns (B) | 1 | 1397.5 | 1397.5 | 363 (1,44) | 0 |
| Interaction AB | 1 | 1.7 | 1.7 | 0.4 (1,44) | 0.5 |
| Error | 44 | 169.4 | 3.9 | | |
| Total | 47 | 1568.8 | 33.4 | | |

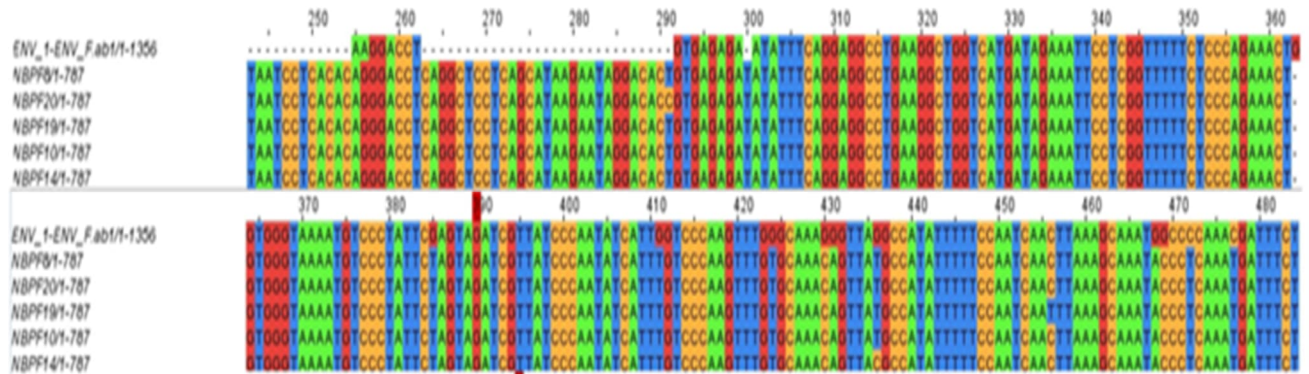
Rys.1.1.2a. Zrzut ekranu ilustrujący wprowadzenie danych do tabeli w dwuczynnikowej analizie wariancji.

- ▶ Czynniki A oznacza płeć. Wartość graniczna dla statystyki F wynosi 1,44, a wartość dla płci wynosi 0,05. Jest więc ona niższa od wartości granicznej. Tym samym nie ma podstaw do odrzucenia hipotezy zerowej o równości średnich dla kobiet i mężczyzn. **Należy przyjąć, że dla obu badanych typów zmian nie wystąpiły różnice między kobietami i mężczyznami.**
- ▶ Wartość F dla interakcji pomiędzy czynnikami, czyli płcią a typem zmiany wywołanej transpozonomem *SIRE-1* wynosi 0,4 i jest niższa od wartości granicznej 1,44. Oznacza to, że nie ma podstaw do odrzucenia hipotezy zerowej o braku interakcji. Oznacza to, że obie płcie podobnie zachowują się pod względem efektów działania *SIRE-1*. **U obu płci częstsze są insercje niż delecje.**
- ▶ Czynniki B to insercje i delecje. Wartość statystyki F wynosi 363 i jest wyższa od wartości granicznej, 1,44. Oznacza to, że należy odrzucić hipotezę o równości średnich. **Średnie dla insercji są różne od średnich dla delecji, przy czym liczba insercji jest wyższa.** Wyższa wartość insercji świadczy o aktywności transpozonu *SIRE-1*.

- Wartości średnich można odczytać z tabeli poniżej obliczeń.

B. Rysunek 1.1.2 przedstawia uliniowanie sekwencji *SIRE1* o długości 242 bp z genami rodziny NBPF. Proszę zidentyfikować wszystkie mutacje, które zaszyły w *SIRE1* w trakcie ewolucji naczelnych. Proszę uwzględnić tylko te mutacje, które występują we wszystkich genach *NBPF*. Na podstawie zidentyfikowanych mutacji proszę podać procent zasad, które uległy zmianie w *NBPF*.

● Mutacje



Rys.1.1.2. Uliniowanie sekwencji transpozonu *SIRE1* oraz genów *NBPF* człowieka.

- ▶ Dwie duże insercje: jedna 12 bp i druga – 29 bp.
- ▶ Insercja T w pozycji 300 bp.
- ▶ Delecja G w pozycji 363 bp.
- ▶ Transwersja G na T w pozycji 385 bp.
- ▶ Transwersja G na T w pozycji 411 bp.
- ▶ Transwersja G na T w pozycji 424 bp.
- ▶ Transwersja G na C w pozycji 430 bp.
- ▶ Tranzycja G na A w pozycji 468 bp.
- ▶ Transwersja G na C w pozycji 469 bp.
- ▶ Tranzycja C na T w pozycji 472 bp.
- ▶ Tranzycja C na T w pozycji 477 bp.

● Łączna liczba zmienionych zasad wynosi 51.

● Procent zasad, które uległy zmianie wynosi $51/242 \times 100\% = 21\%$

1.2. Identyfikacja sekwencji transpozonowych

1.2.3. Która para sekwencji DNA może zostać zakwalifikowana jako TIR przy założeniu, że przedstawiają one sekwencję na tej samej nici DNA?

- A. 5'GAATCCGCA3' i 5'ACGCCTAAG3'
- B. 5'GAATCCGCA3' i 5'CTTAGGCGT3'
- C. 5'GAATCCGCA3' i 5'GAATCCGCA3'
- D. 5'GAATCCGCA3' i 5'TGCGGATTC3'**

● **Jest to sekwencja D.** Jeżeli uwzględniamy jedną nić, to sekwencja czytana w kierunku 5'-3' musi być komplementarna to sekwencji czytanej w kierunku 3'-5', czyli powtórzenie ma sekwencję: 5'TGCGGATTC3'. Warunek ten spełnia sekwencja D.

1.2.4. Która para sekwencji wskazuje na duplikację w miejscu insercji przy założeniu, że przedstawiają one sekwencję na tej samej nici DNA? Proszę wyjaśnić.

- A. 5'AATTCGCGT3' i 5'AATTCGCGT3'
- B. 5'AATTCGCGT3' i 5'TGCGCTTAA3'
- C. 5'AATTCGCGT3' i 5'TTAAGCCGCA'
- D. 5'AATTCGCGT3' i 5'ACGCGAATT3'

- Jest to para A. Duplikacja w miejscu insercji oznacza powtórzenie tej samej sekwencji w obrębie jednej nici. Wymóg ten spełnia sekwencja A.

3. Aktywność transpozonów u człowieka

3.2. Choroby człowieka wywołane insercją transpozonów

3.2.1. Hemofilia typu A jako efekt insercji *Alu* do genu kodującego czynnik FVIII.

Jedną z przyczyn hemofilii typu A jest insercja transpozonu *Alu* do genu kodującego czynnik FVIII. Częstość insercji transpozonu *Alu* w gametach wynosi 5%. Częstość występowania hemofilii A u mężczyzn w Polsce wynosi 1/5000. Gen kodujący czynnik VIII ma długość 186 kbp. Czy obserwowaną częstość hemofilii A u mężczyzn w Polsce można wytłumaczyć tylko nowymi insercjami transpozonu *Alu*? Proszę uzasadnić przyjmując, że prawdopodobieństwo insercji jest identyczne dla całego genomu człowieka.

- Należy oszacować jak często wystąpi hemofilia związana z insercją *de novo* transpozonu *Alu*. Częstość insercji *Alu* w dowolny region genomu wynosi 5% (0,05), czyli 5% gamet ma insercję *Alu* w dowolnym regionie genomu.
- Gen kodujący FVIII ma 186 kbp i stanowi $186 \times 10^3 / 3,2 \times 10^9$ genomu człowieka, co daje w przybliżeniu prawdopodobieństwo wystąpienia genu kodującego FVIII - $0,58 \times 10^{-4}$. Prawdopodobieństwo insercji do genu FVIII podczas tworzenia gamet u matki wynosi $0,58 \times 10^{-4} \times 0,05$, co daje w przybliżeniu 3×10^{-6} . Gen FVIII zlokalizowany jest w chromosomie X, który mężczyźni otrzymują od matki, częstość insercji do genu FVIII u matki odpowiada częstości wystąpienia gamet z insercją u matki. Prawdopodobieństwo, że gameta z insercją połączy się z gametą zawierającą Y wynosi 0,5. Zatem częstość wystąpienia insercji u mężczyzn wynosi $3 \times 10^{-6} \times 0,5 = 1,5 \times 10^{-6}$. Zatem częstość wystąpienia hemofilii u mężczyzn na skutek insercji *de novo* w gametach matki wynosi 1,5/1 mln.
- Częstość występowania hemofilii u mężczyzn w Polsce to 1/5000, czyli 200 na mln. Częstość występowania hemofilii u mężczyzn w Polsce jest około 155 razy wyższa niż wynikałoby to z częstości insercji *Alu*. Nowe insercje stanowią około 0,75% wszystkich przypadków hemofilii w Polsce.