

Ćwiczenie 14

Metody analizy transpozonów. Identyfikacja sekwencji transpozonowych. Transpozony w stanach chorobowych.

Kornelia Polok

1. Metody analizy transpozonów

➔ Transpozony występują powszechnie w genomach Eukariota, w tym w genomie człowieka. 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. Metody molekularne identyfikacji transpozonów

1.1.1. Transpozony należą do rodzin, które charakteryzują się określonymi motywami strukturalnymi, które można wykorzystać do przeszukiwania sekwencji genomowych oraz do projektowania starterów. Projektowanie starterów komplementarnych do fragmentu DNA pomiędzy transpozonom, a miejscem restrykcyjnym w sekwencji powtarzalnej jest skuteczną metodą identyfikacji miejsc insercji transpozonów w genomie. Jeżeli transpozon integruje się w nowym miejscu, jest to obserwowane jako polimorfizm (Rys. 1.1.a).

● Genomowy DNA jest trawiony enzymami restrykcyjnymi:

- ▶ MseI, który rozpoznaje sekwencję: TTAA;
- ▶ PstI, który rozpoznaje sekwencję CTCGAG.



● W wyniku cięcia enzymami restrykcyjnymi otrzymujemy fragmenty, które zawierają miejsce restrykcyjne oraz transpozon.

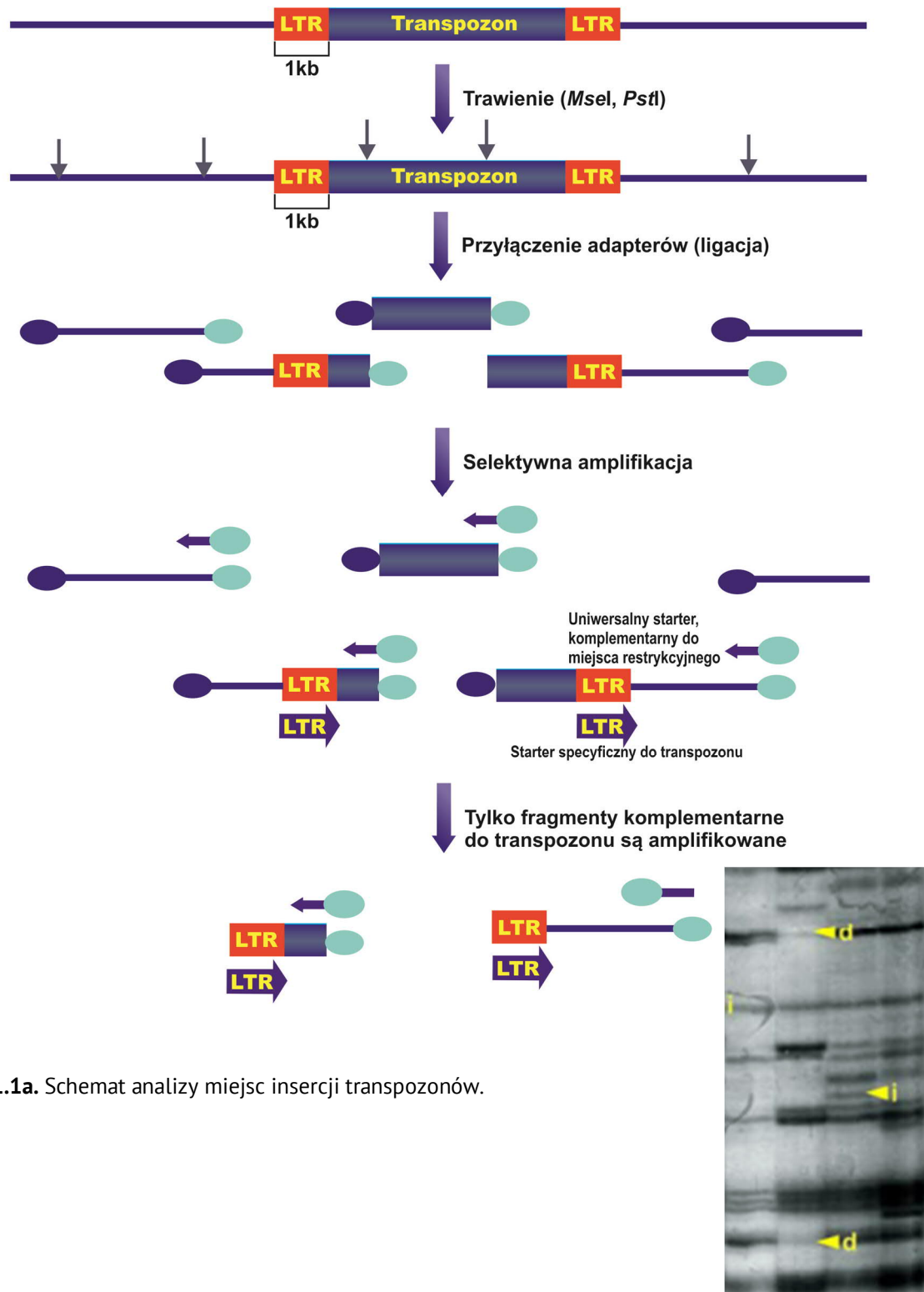


● Otrzymane fragmenty mają lepkie końce. Przyłączamy do nich adaptory o ustalonej sekwencji nukleotydowej oraz komplementarne do lepkich końców powstałych w wyniku cięcia enzymami restrykcyjnymi.

● Dodajemy startery:

- ▶ Jeden starter jest komplementarny do adaptera. Dodatkowo wprowadza się 1-3 losowe zasady, aby uwzględnić polimorfizm w miejscu sekwencji.

- ▶ Drugi starter jest komplementarny do sekwencji transpozonu,
- Dzięki doborowi dwóch starterów amplifikowany będzie fragment zawierający transpozon oraz sekwencję miejsca insercji.
- 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 delekcję.



Rys.1.1a. Schemat analizy miejsc insercji transpozonów.



1.1.2. Pewną linię komórkową napromieniowano, a następnie porównano miejsca insercji transpozonu LTR w zmutowanych komórkach oraz kontroli. W kontroli stwierdzono 89 insercji, natomiast wyniki dla napromieniowanych linii podano w tabeli 1.

- A. Na podstawie podanych wyników proszę ocenić częstość insercji oraz delecji. **(2 punkty)**
 B. Czy w napromieniowanych komórkach doszło do remobilizacji transpozonu LTR? Proszę uzasadnić odpowiedź. **(1 punkt)**

Tabela 1. Porównanie miejsc insercji w liniach komórkowych po napromieniowaniu.		
Linia	Nowe insercje	Delecje
L1	9	6
L2	4	2
L3	12	4
L4	11	5
L5	16	8
L6	20	8
L7	17	9
L8	8	3

1.1.3. Delecja miejsc insercji często jest wynikiem powstawania tzw. solo-LTR. Jak mógł powstać solo LTR? **(1 punkt)**

1.1.4. Transpozon THE często występuje w sekwencjach telomerowych. Jak udowodnić jego lokalizację w telomerach? **(1 punkt)**

1.2. Metody informatyczne



Znajdują zastosowanie, gdy nie posiadamy wiedzy o sekwencji transpozonu. 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 lokalizacja w innych powtórzeniach;
- fragmentacji sekwencji
- rozdzieleniem blisko spokrewnionych rodzin, które mają wspólną znaczną część sekwencji.

1.2.1. Etapy analizy *de novo*

- 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 przerywane insercjami o długości >200 bp. Metoda użyteczna przy wykrywaniu nowych rodzin transpozonowych.

2. Identyfikacja sekwencji transpozonowych

2.1. Odwrócone powtórzenia terminalne (TIR)

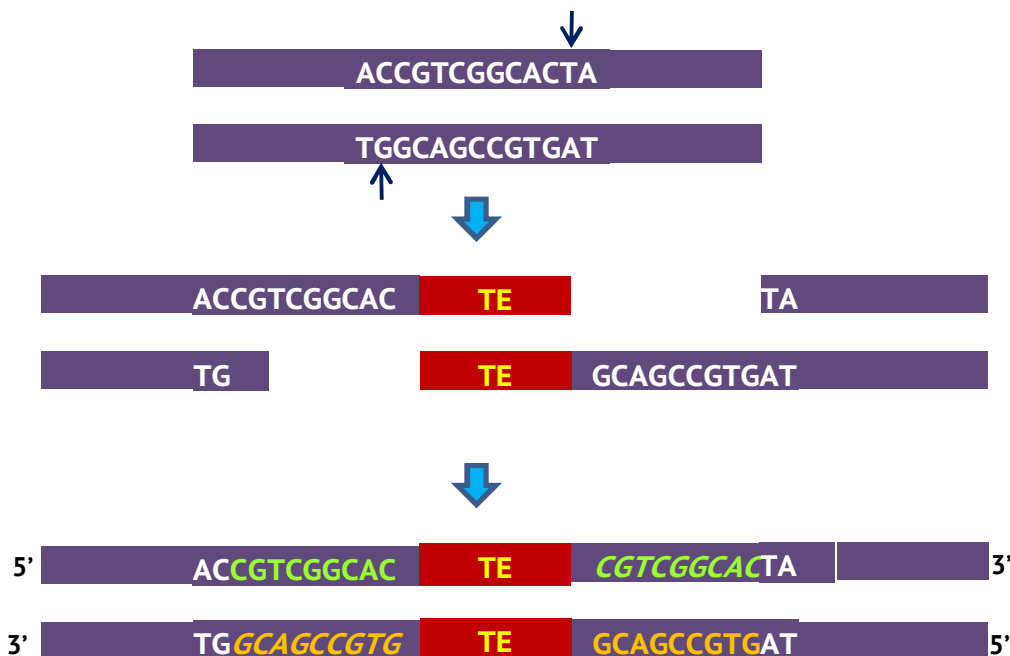
➔ Sekwencje zlokalizowane na końcach transpozonów większości transpozonów DNA u Eukariota. Występują także w elementach insercyjnych Prokariota. Elementy insercyjne są najprostszymi transpozonami, które zawierają tylko geny niezbędne do przemieszczania. Jeżeli dwa elementy insercyjne są blisko siebie,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



Rys.1.3.1. Odwrócone powtórzenia (TIR) na końcach elementów insercyjnych.

2.2. Duplikacje w miejscu insercji

Gdy transpozon wbudowuje się do genomu to w miejscu insercji dochodzi do duplikacji sekwencji w miejscu docelowym. Sekwencje te mają długość 2-13 bp. Flankują one transpozon. Prawdopodobnie są efektem asymetrycznego trawienia DNA donorowego podczas wbudowywania.



Rys.1.3.2. Powstawanie duplikacji w miejscu insercji transpozonu. Kursywą zaznaczono dosyntetyzowaną sekwencję. Kolor oznacza powtórzenie.

2.3. Zadania

2.3.1. Która para sekwencji DNA może zostać zakwalifikowana jako TIR przy założeniu, że przedstawiają one sekwencję na tej samej nici DNA? Proszę wyjaśnić. **(1 punkt)**



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

2.3.2. Która para sekwencji wskazuje na duplikację w miejscu insercji przy założeniu, że przedstawiają one sekwencję na tej samej nici DNA? Proszę wyjaśnić/ **(1 punkt)**

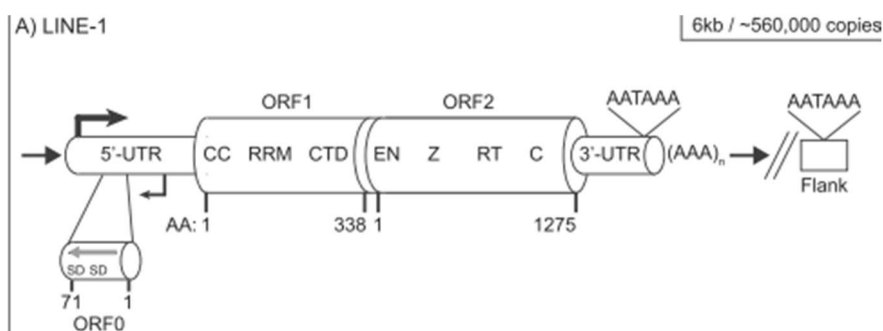
- A. 5'AATTCGCGT3' i 5'AATTCGCGT3'
- B. 5'AATTCGCGT3' i 5'TGCGCTTAA3'
- C. 5'AATTCGCGT3' i 5'TTAAGCCGCA'
- D. 5'AATTCGCGT3' i 5'ACGCGAATT3'

3. Transpozony w stanach chorobowych.

3.1. Aktywność transpozonu LINE-1

➔ Niektóre transpozony są nadal aktywne u człowieka. Dotyczy to około 80-100 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 także łańcuch poliadenylowy.

- ORF1p: białko o masie 40 kDa, 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.
- ORF2p: białko o masie 150 kDa, aktywność endonukleaz i odwrotnej transkryptazy.



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

- LINE1 jest transkrybowany z własnego promotora zlokalizowanego w 900 bp fragmencie 5'UTR przez polimerazę RNA II. Ponadto w regionie 5'UTR znajduje się 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 200 bp regionu UTR3'
- W wyniku transkrypcji LINE1, RNA jest transportowany do cytoplazmy, gdzie dochodzi do translacji ORF1p i ORF2p i złożenia transpozozonu.
- Złożony LINE1 składa się z trimerów ORF1p, ORF2p, LINE1-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, która 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, RNA kodowanego przez genom – U6 RNA oraz mRNA dla białek. Genom człowieka zawiera około 8 tys. pseudogenów, które są retrokopiami i obecnie podlegają obróbce.

3.2. Choroby wywołane insercją retrotranspozonów

Zidentyfikowano około 124 jednostki chorobowe na skutek insercji transpozozonu 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 transpozozonu może spowodować zmianę ekspresji genu. Dane z sekwencjonowania genomów, zwłaszcza sekwencjonowania pojedynczych komórek wskazują, że insercje w somatycznych komórkach mogą być częste.

- Insercja 1722 bp do genu *CYBB*, mukowiscydoza, *LINE1*
- Insercja do genu *BRCA1*, dziedziczny nowotwór piersi, *Alu*.
- Insercja 2524 bp do genu *FIX*, hemofilia typu B, *SVA*.

3.3. Hemofilia jako efekt insercji *Alu* do genu *FVIII*



Jedną z przyczyn hemofilii typu A jest insercja transpozozonu *Alu* do genu *FVIII*. Częstość insercji transpozozonu *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 transpozozonu *Alu*.

(2 punkty)

Odpowiedzi

1. Metody analizy transpozonów

1.1. Metody molekularne

1.1.2. Pewną linię komórkową napromieniowano, a następnie porównano miejsca insercji transpozonu LTR w zmutowanych komórkach oraz kontroli. W kontroli stwierdzono 89 insercji, natomiast wyniki dla napromieniowanych linii podano w tabeli 1. Na podstawie podanych wyników proszę ocenić częstość insercji oraz delecji. Czy w napromieniowanych komórkach doszło do remobilizacji transpozonu LTR? Proszę uzasadnić odpowiedź.

Tabela 1. Porównanie miejsc insercji w liniach komórkowych po napromieniowaniu.		
Linia	Nowe insercje n (%)	Delecje n (%)
L1	9 (10,1)	6 (6,7)
L2	4 (4,5)	2 (2,2)
L3	12 (13,5)	4 (4,5)
L4	11 (12,4)	5 (5,6)
L5	16 (18,0)	8 (9,0)
L6	20 (22,5)	8 (9,0)
L7	17 (19,1)	9 (10,1)
L8	8 (9,0)	3 (3,4)

- Należy obliczyć procent insercji i delecji w każdej linii uwzględniając liczbę insercji w kontroli. Częstość insercji wynosiła 4,5 do 22,5%, natomiast częstość delecji – 2,2 do 10%.
- Porównanie analizowanych komórek wskazuje na dwukrotnie wyższą częstość insercji niż delecji. Oznacza to, że mogła nastąpić remobilizacja analizowanego transpozonu.

1.1.3. Delecja miejsc insercji często jest wynikiem powstawania tzw. solo-LTR. Jak mógł powstać solo LTR? **(1 punkt)**

- Solo-LTR powstają na skutek crossing-over między sekwencjami LTR transpozonu. Mechanizm ten jest uważany za obronę organizmu przed nadmierną mobilizacją transpozonów.

1.1.4. Transpozon THE często występuje w sekwencjach telomerowych. Jak udowodnić jego lokalizację w telomerach? **(1 punkt)**

- Należy wykorzystać hybrydyzację *in situ*, przy czym sondą powinna być sekwencja transpozonu THE

2. Identyfikacja sekwencji transpozonowych

2.3. Zadania

2.3.1. Która para sekwencji DNA może zostać zakwalifikowana jako TIR przy założeniu, że przedstawiają one sekwencję na tej samej nici DNA? (1 punkt)

- 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 uwzględnimy 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.

2.3.2. Która para sekwencji wskazuje na duplikację w miejscu insercji przy założeniu, że przedstawiają one sekwencję na tej samej nici DNA? Proszę wyjaśnić/ (1 punkt)

- 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. Transpozony w stanach chorobowych

3.1. Hemofilia jako efekt insercji *Alu* do genu *FVIII*

Jedną z przyczyn hemofilii typu A jest insercja transpozonu *Alu* do genu *FVIII*. Częstość insercji transpozonu *Alu* w gametach wynosi 5%. Częstość występowania hemofilii A u mężczyzn w Polsce wynosi 1/5000. Czy obserwowaną częstość hemofilii A u mężczyzn w Polsce można wytłumaczyć tylko nowymi insercjami transpozonu *Alu*.

- 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 *FVIII* ma 186 kbp i stanowi $186 \times 10^{-3} / 3,2 \times 10^9$ część genomu człowieka, co daje w przybliżeniu $0,58 \times 10^{-4}$.
- 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} . Ponieważ *FVIII* zlokalizowany jest w chromosomie X, który mężczyźni otrzymują od matki, częstość insercji do genu *FVIII* odpowiada częstości wystąpienia tej choroby u mężczyzn. Zatem częstość wystąpienia hemofilii u mężczyzn na skutek insercji *de novo* w gametach matki wynosi 3/1 mln.
- Częstość występowania hemofilii u mężczyzn w Polsce to 1/5000 czyli 20 na 100 tys. Częstość występowania hemofilii u mężczyzn w Polsce jest około 7 razy wyższa niż wynikałoby to z częstości insercji *Alu*. Nowe insercje stanowią około 1,5% wszystkich przypadków hemofilii w Polsce.