

Recenzja
rozprawy doktorskiej mgr Magdaleny Reginy Kubiak pt.
„*Udział retrokopii genów kodujących białka w kształtowaniu genomu, transkryptomu*
***i proteomu człowieka*”**
[Contribution of retrocopies derived from protein-coding genes to shaping human
genome, transcriptome and proteome]

Prowadzone przez minione dwadzieścia lat badania zmieniły pierwotny pogląd na kopie genów powstające w wyniku retropozycji, tj. odwrotnej transkrypcji mRNA i wbudowania cDNA w sekwencję genomową. Retrokopie, do niedawna uznawane za niefunkcjonalne pseudogeny stanowiące tzw. śmieciowe DNA, stały się w ostatnich latach przedmiotem intensywnych badań i analiz. Coraz więcej doniesień wskazuje na to, że przynajmniej część z retrokopii genów kodujących białka ulega ekspresji. Praca doktorska mgr Magdaleny Kubiak doskonale wpisuje się w nurt badań nad rolą takich funkcjonalnych retrokopii w kształtowaniu genomu, transkryptomu i proteomu człowieka.

Przedłożona do oceny dysertacja jest, w rozumieniu nowej procedury przewodu doktorskiego, osiągnięciem naukowym, na które składa się spójny tematycznie zbiór pięciu anglojęzycznych artykułów opublikowanych w czasopismach naukowych. Praca powstała przy wsparciu finansowym uzyskanym z NCNu (grant Preludium), Poznańskiego Konsorcjum RNA KNOW oraz programu Power (interdyscyplinarne studia doktoranckie).

W dysertacji przedstawione zostały wymagane ustawowo oświadczenia współautorów, określające ich rolę w przygotowaniu publikacji stanowiących podstawę osiągnięcia naukowego. Pozycja mgr Kubiak jako pierwszego lub drugiego autora w czterech z tych prac, jak również opis udziału Doktorantki w badaniach i redagowaniu publikacji pozwala wnioskować, że Jej wkład w powstaniu osiągnięcia naukowego stanowiącego podstawę dysertacji był znaczący. Dwie z publikacji to prace o charakterze eksperymentalnym, jedna oparta jest o szeroko zakrojoną analizę bioinformatyczną; pozostałe dwie publikacje to prace przeglądowe.

Dołączony do pięciu publikacji tekst obejmuje 12-stronicowe omówienie tematu w języku polskim i angielskim, opisujące pokrótce stan wiedzy w zakresie tematu pracy oraz opublikowane wyniki i plany badawcze. Dołączony do wprowadzenia spis publikacji liczy 29 pozycji; pełna lista literatury w zakresie omawianych zagadnień znajduje się w poszczególnych publikacjach.

W tekście wprowadzającym Autorka omówiła proces retropozycji, zwracając uwagę na charakter i potencjał aktywnych retrokopii, zwanych dalej retrogenami. Przedstawiła eksperymentalne metody identyfikacji retrokopii i metody weryfikacji ekspresji tych sekwencji opisanych w unikatowej bazie adnotacji retrokopii RetrogeneDB. Baza ta stanowi punkt

odniesienia (zarówno w kontekście wyboru badanych sekwencji jak i deponowania uzyskanych wyników) dla działań eksperymentalnych wykonanych w ramach pracy doktorskiej. W doktoracie Autorka zajęła się trzema tematami, które pokrótce przedstawione zostały we wprowadzeniu.

Pierwszy temat to badanie wpływu retropozycji genów na kształtowanie różnic międzygatunkowych. W oparciu o analizę danych RNAseq i EST Doktorantka przeprowadziła eksperymentalne badania ekspresji retrokopii zidentyfikowanych we wcześniejszych pracach Zespołu. Badania te potwierdziły predykcje bioinformatyczne w odniesieniu do 62 retrokopii adnotowanych w genomie człowieka lub myszy i ukazały potencjalną aktywność sekwencji specyficznych gatunkowo oraz przyczyniły się do rozszerzenia bazy danych. Badania ekspresji opisane są w publikacji w czasopiśmie *Database – Oxford* (2017), gdzie Doktorantka jest umieszczona w zespole autorskim na piątym miejscu (15%).

Drugim celem badawczym była analiza zmienności retroduplikacji w populacjach ludzkich. Doktorantka i jej koledzy, analizując dane z projektu Sekwencjonowania 1000 Genomów, po raz pierwszy wykazali, że międzypopulacyjna zmienność w zakresie retroduplikacji może być wynikiem nie tylko insercji, ale także utraty retrokopii. Rola Doktorantki polegała na eksperymentalnej weryfikacji obecności pięciu polimorficznych sekwencji w różnych populacjach ludzkich oraz na wykonaniu szczegółowej analizy porównawczej genomów człowieka, szympansa i goryla. Analiza ta wykazała, że źródłem polimorfizmu są delecje retrokopii w genomie człowieka. W dalszej części pracy Autorka, opierając się na danych Zespołu badawczego, przeprowadziła eksperymentalną walidację kolejnych 193 retrokopii, które zostały utracone w niektórych populacjach. Wyniki zostały opublikowane w czasopiśmie *PLOS Genetics* (2015), gdzie Doktorantka jest drugim autorem z udziałem 20%.

Ostatnim, głównym elementem rozprawy, była kompleksowa analiza bioinformatyczna funkcji pełnionych przez retrokopie człowieka. W ramach tej analizy Doktorantka wraz z jej Promotorką (prof. Makałowską) opublikowała, jako pierwszy autor z udziałem 80%, pracę przeglądową (*Viruses*, 2017), w której przedstawiła stan wiedzy na temat funkcji retrokopii kodujących białka. W dalszej, najbardziej rozbudowanej części rozprawy, Doktorantka wraz z Zespołem dokonała bioinformatycznej charakterystyki funkcji pełnionych w genomie, transkryptomie i proteomie człowieka przez retrokopie zdeponowane w bazie RetrogeneDB. Wyniki dotyczące tej kompleksowej analizy, na podstawie której dla części retrokopii zaproponowano funkcje regulatorowe oraz wytypowano kandydatów do badań mechanizmów w sieciach interakcji pomiędzy retrokopiami i genami rodzicielskimi, opisane zostały w pracy opublikowanej w *Genes* (2020), w której Doktorantka jest pierwszym autorem z udziałem 70%. Prace w tej części dysertacji zamyka kolejna publikacja przeglądowa (*RNA*, 2018), w której

opisano molekularne podstawy działania naturalnych transkryptów antysensownych (w tym tych pochodzących z retrokopii) w chorobach człowieka; Doktorantka jest tu drugim autorem z udziałem 25%.

Ocena publikacji przedstawiona jest w kolejności ich omówienia w dysertacji.

Publikacja w *Database-Oxford*, 2017. Praca ta poświęcona jest opisowi organizacji, rozbudowy i aktualizacji bazy danych retrokopii (RETROGENEDB). Baza zawiera adnotacje ponad 86 tys. retrokopii w 62 zwierzęcych i 37 roślinnych genomach, w tym informacje na temat konserwacji sekwencji, sprzężonej aktywności polimerazy oraz obecności alternatywnych miejsc startu transkrypcji (TSS), a także ortologicznych relacji pomiędzy retrogenami. Publikacja jest rzetelnym opisem narzędzi dotyczących adnotowania sekwencji, jednak nie dotyczy badania wpływu retropozycji genów na kształtowanie różnic międzygatunkowych, czego można by się spodziewać na podstawie opisu zamieszczonego w streszczeniu i wstępie do dysertacji. Ta niespójność nie zmienia wysokiej oceny wartości i unikatowego charakteru bazy RETROGENE, a to między innymi ze względu na zamieszczoną w niej, potwierdzoną eksperymentalnie przez mgr Kubiak, informację na temat ekspresji retrokopii. Poniżej wymieniam kilka drobnych zastrzeżeń:

- W opisie identyfikacji retrokopii znajduje się stwierdzenie, że przeszukiwano sekwencje genów wieloeksonowych, zdefiniowanych jako posiadające dwa lub więcej eksonów; dalej znajduje się wskazanie, że alignment musiał obejmować co najmniej dwa rodzicielskie introny – czyli geny wieloeksonowe musiały posiadać przynajmniej trzy eksony?

- W pracy poruszona jest kwestia trudności eksperymentalnej identyfikacji ekspresji retrokopii metodą PCR, jednak brak jest informacji, w jaki sposób poradzono sobie z tym problemem.

- W Tabeli ST3 brak wyjaśnienia skrótu TSI opisującego jedną z populacji.

Publikacja w *PLOS Genetics* (2015) jest opisem badania międzypopulacyjnych różnic ekspresji związanych z polimorficzną utratą retrokopii. Analiza zmienności wynikającej z utraty ancestralnych retrokopii (w przeciwieństwie do szeroko badanych niedawnych retroduplikacji) jest nowym podejściem. Używając danych z Projektu 100 Genomów Autorzy zbadali populacyjną zmienność występowania około 5 tys. retrokopii opisanych w RETROGENEDB. Wykryto 193 przypadki RDV, w większości wynikające z utraty retrokopii ancestralnej. W celu rozróżnienia pomiędzy RDV wynikającymi z delecji i insercji Doktorantka przeprowadziła analizę dywergencji sekwencji retrokopii oraz występowania ich ortologów w innych gatunkach; na tej podstawie potwierdziła m.in. wzrost intensywności retropozycji u Naczelnych, a także zidentyfikowała nowe retrokopie, których polimorfizm u człowieka jest wynikiem delecji ancestralnych retrokopii. Wykorzystując dostępne dane RNAseq, przeprowadziła dodatkowo

analizę ekspresji retrokopii w różnych populacjach ludzkich. Ekspresja 11 ancestralnych retrokopii, które u niektórych osób uległy delecji, została potwierdzona w różnych tkankach; ponadto, stwierdzono statystycznie istotne różnice w poziomie ekspresji niektórych retrogenów w populacjach. Dodać należy, że opis metod stosowanych w celu uzyskania specyficznych wyników PCR, zakwestionowany w poprzedniej publikacji, tutaj jest w pełni satysfakcjonujący; ponadto, wykonane analizy zostały zilustrowane odpowiednimi rycinami i schematami.

Praca przeglądowa [*Viruses*, 2017 r.], w której Doktorantka jest pierwszym autorem, stanowi rzetelne omówienie stanu wiedzy w zakresie retrokopii genów kodujących białka i ich funkcji, oparte o aktualną w 2017 roku i trafnie dobraną literaturę (180 pozycji). W omawianej pracy przeglądowej mgr Kubiak wykazała dobrą znajomość tematu funkcjonalnych retrogenów i umiejętność logicznego przedstawienia złożonych zagadnień.

We pierwszym rozdziale omówione zostały retropozony jako źródło odwrotnej transkryptazy w komórce i różne związane z tym mechanizmy retropozycji, a w kolejnym – informacje dotyczące liczby retrokopii w różnych genomach. W dalszej części Autorka opisała molekularne funkcje pełnione przez retrokopie genów, z podziałem na mechanizmy działania na poziomie DNA, RNA i białka; na pochwałę zasługują schematyczne ryciny, bardzo trafnie podsumowujące istotę tych mechanizmów. W osobnych podrozdziałach przedstawione są: przykłady ewolucji retrokopii kodujących białka w genomach różnych organizmów związane z subfunkcjonalizacją, nabywaniem nowej funkcji czy przejęciem funkcji genu rodzicielskiego, konsekwencje insercji retrokopii do genu docelowego, czy rola retrokopii w promowaniu homologicznej rekombinacji. Osobna część pracy poświęcona jest przedstawieniu roli i mechanizmu działania retrotranskryptów w regulacji ekspresji różnych genów. Pracę kończą dwa krótkie rozdziały, poświęcone roli retrogenów w patogenezie chorób oraz wykorzystaniu retrokopii jako źródła informacji o różnorodności genomu w populacjach.

Publikacja w *Genes*, 2020. Praca ta jest opisem głównej części dysertacji i dotyczy kompleksowego badania bioinformatycznego mającego na celu scharakteryzowanie występowania i funkcji retrokopii genów kodujących białka. We wstępie mgr Kubiak podkreśliła, jak ważne jest badanie funkcjonalnych retrokopii w świetle dokonanych w ostatnim dziesięcioleciu obserwacji, że większość RNA nie koduje białek, a niekodujące RNA pełnią niezmiernie ważną funkcję w regulacji funkcjonowania komórek. Mimo rosnącego zainteresowania badaczy tym zagadnieniem, w istniejącej wiedzy istnieją duże luki, stąd zaangażowanie zespołu prof. Makołowskiej w badania tej funkcji jest w pełni uzasadnione. W publikacji przedstawiającej kompleksową analizę udziału retrogenów w kształtowaniu i ekspresji genomu człowieka, Doktorantka występuje w roli wiodącej. Korzystając z bogactwa dostępnych danych będącego wynikiem rozwoju metod wysokoprzepustowej analizy genomów, transkryptomów i epigenomów różnych organizmów, przeprowadziła analizę wachlarza

możliwych funkcji pełnionych przez obecne w ludzkim genomie retrokopie genów kodujących białka. Publikacja zawiera przejrzysty opis stosowanych w analizie bioinformatycznej metod; w pełni uzasadniony dobór metod jest jednym z licznych dowodów kompetencji zespołu autorskiego.

W omawianej pracy wyróżniono trzy poziomy aktywności retrokopii. Pierwszy to kodowanie białek. Autorka zidentyfikowała 117 pozornie niekodujących retrokopii, które mogą ulegać translacji, w tym 14 retrokopii wykazujących ekspresję w większości tkanek. Nowym odkryciem było stwierdzenie wysokiej ekspresji retrogenów w śledzionie; porównanie ekspresji w normalnych i rakowych tkankach pozwoliło też zidentyfikować trzy retrogeny ulegające ekspresji we wszystkich liniach nowotworowych, ale nie w tkankach normalnych. Analiza GO wskazała na kategorię złączy komórkowych przypisaną do retrogenu ARF6, retrokopii genu SRF, który takiej funkcji nie posiada; ponadto, znaleziono kilka par gen-retrogen różniących się w nieznacznym stopniu funkcją i/lub subkomorkową lokalizacją, sugerującą możliwość sub- lub neofunkcjonalizacji domen. Na podstawie analizy danych proteomicznych zidentyfikowano 740 retrogenów i odpowiadających im unikatowo sekwencji peptydów. Zidentyfikowano 71 transkryptów genów białkowych zawierających sekwencje retrokopii; poza retrokopiami wbudowanymi regiony UTR, znaleziono 18 retrokopii, których sekwencja uległa eksaptacji jako nowe eksony w genach kodujących białek.

Drugi poziom aktywności retrokopii to regulatorowa funkcja ich transkryptów. Transkrypty te mogą działać jako kompetytywne endogenne RNA (gąbki miRNA); ich korelacja z ekspresją regulowanych genów jest na ogół pozytywna, zidentyfikowano jednak przypadki korelacji negatywnej. Wiele retrogenów ulega transkrypcji jako długie niekodujące RNA; transkrypty te mogą działać regulatorowo zarówno w stosunku do genów rodzicielskich w trans jak i docelowych w cis. Doktorantka wyodrębniła 181 transkryptów mogących funkcjonować jako kompetytywne RNA oraz liczne potencjalne transkrypty działające regulatorowo wobec genów rodzicielskich (67) oraz w cis (656). W publikacji przedstawiony został szczegółowo tok przeprowadzonych analiz oraz omówione ciekawe przykłady retrokopii pełniących wymienione funkcje.

Trzeci poziom aktywności związany jest z obecnością retrokopii w DNA. Na podstawie analizy korelacji poziomu ekspresji genów docelowych z obecnością zagnieżdżonych w nich retrokopii Doktorantka wyodrębniła 17 przypadków interferencji transkrypcyjnej, wpływającej na splicing genów, w których ulokowane są retrokopie, a na podstawie analizy transkryptów fuzyjnych zidentyfikowała 15 retrokopii związanych z procesami rekombinacji.

W ostatniej części pracy przedstawione zostały przykłady sieci regulatorowych, wskazujące na to, że retrokopie mogą pełnić więcej niż jedną funkcję, a regulacja może być zarówno pozytywna i negatywna. We wnioskach zawartych w publikacji znalazło się bardzo

śluszne stwierdzenie, że w związku z powszechną aktywnością transkrypcyjną setek retrokopii i różnorodnością mechanizmów ich wpływu na ludzki genom, zakonserwowanie otwartej ramki odczytu czy obecność promotora, nie są – jak do niedawna uważano – jedynym kryterium określenia funkcjonalności retrokopii. Reasumując, przedstawiona publikacja stanowi opis naprawdę kompleksowej analizy i może służyć jako encyklopedia wiedzy o roli funkcjonalnych retrokopii i mechanizmach ich wpływu na kształtowanie i ekspresję ludzkiego genomu.

Praca przeglądowa w RNA, 2017, stanowi solidne i bardzo przydatne podsumowanie wiedzy dotyczącej naturalnych transkryptów antysensownych w patogenezie różnych chorób i ich potencjalnego wykorzystania w terapii. Przedstawione informacje oparte są o aktualną i trafnie dobraną literaturę (150 pozycji). Omówienie mechanizmów działania transkryptów antysensownych obejmuje znacznie szerszy zakres niż ten dotyczący jedynie retrokopii przedstawiony w poprzednich publikacjach (np. modyfikacja histonów, metylacja wysp CpG, imprinting, czy mechanizmy związane z siRNA). Poza szczegółowym przedstawieniem molekularnych podstaw zaangażowania naturalnych transkryptów antysensownych w etiologię różnych chorób, jak nowotwory, choroby neurodegeneracyjne czy dysfunkcje układu krążenia, w pracy przedyskutowano plusy i minusy potencjalnego zastosowania terapii opartych na użyciu oligonukleotydów skierowanych przeciw transkryptom antysensownym.

W mojej ocenie, dysertacja i stanowiące jej podstawę publikacje posiadają bardzo wysoką wartość poznawczą i praktyczną. Wyniki badań, które mgr Kubiak przeprowadziła wraz z zespołem badawczym prof. Makałowskiej, demonstrują międzygatunkową i międzypopulacyjną zmienność retropozycji wskazującą na rolę tych elementów w ewolucji, a co więcej, wskazują na różnorodne funkcje, jakie może pełnić znaczna część retrokopii adnotowanych w genomie człowieka. Baza retrokopii, w udoskonaleniu której Doktorantka brała aktywny udział, jest niezwykle cennym narzędziem badawczym i źródłem informacji dla wielu naukowców zajmujących się różnymi zagadnieniami związanymi z retropozycją. Wreszcie, aspekt retrokopii działających jako naturalnych transkryptów antysensownych jest ważkim elementem zrozumienia patogenyzy chorób związanych z zaburzeniami regulacji transkrypcji i translacji.

Znacząca rola mgr Kubiak w wykonaniu części eksperymentalnej/bioinformatycznej jak i w powstawaniu publikacji nie podlega dyskusji; z zadowoleniem odnotowuję, że Doktorantka zamierza kontynuować prace dotyczące retrogenów, w ramach projektu dotyczącego regulacji alternatywnego składania eksonów przez retrokopie znajdujące się w intronach innych genów. Przedstawione przeze mnie nieliczne uwagi krytyczne w niczym nie umniejszają wysokiej oceny wartości dysertacji, a są jedynie wynikiem zainteresowania ciekawym tematem pracy.

Podsumowując, praca mgr Magdaleny Kubiak wnosi nowe, oryginalne elementy i istotnie poszerza wiedzę w zakresie roli funkcjonalnych retropozonów w genomie człowieka, a wkład Doktorantki w planowanie, analizę i opracowanie przedstawionych badań jest znaczący. W związku z powyższym stwierdzam, że przedstawiona do oceny praca spełnia wszystkie wymagania stawiane rozprawom doktorskim. Zwracam się więc do Wysokiej Rady Wydziału Biologii Uniwersytetu Adama Mickiewicza w Poznaniu z wnioskiem o dopuszczenie pani mgr Magdaleny Reginy Kubiak do dalszych etapów przewodu doktorskiego. Równocześnie, biorąc pod uwagę wysoki poziom przedstawionych badań, wnoszę o wyróżnienie przedłożonej do oceny pracy doktorskiej.



Ewa Ziętkiewicz, prof. dr hab.

Instytut Genetyki Człowieka PAN w Poznaniu

Poznań, 28 sierpnia 2020 r.