

POLISH ACADEMY OF SCIENCES



INSTITUTE OF BIOORGANIC CHEMISTRY

Noskowskiego Str. 12/14, 61-704 Poznań, Poland
tel.: +48-61 operator 852 85 03, secretariat 852 89 19
fax: +48-61 852 05 32, e-mail: ibch@ibch.poznan.pl
identifikator 000849327

Poznań, 22 sierpnia 2017

Dr hab. Piotr Kozłowski, Prof. IChB PAN

RECENZJA

rozprawy doktorskiej mgr Wojciecha Rosikiewicza

“Identyfikacja i analiza ekspresji nakładających się genów u człowieka i myszy”

Rozprawa doktorska została przygotowana w Zakładzie Genomiki Zintegrowanej na Wydziale Biologii Uniwersytetu im. Adama Mickiewicza (UAM) w Poznaniu pod kierunkiem prof. dr hab. Izabeli Makałowskiej.

Przedstawiona do recenzji rozprawa doktorska podejmuje ciekawe zagadnienie, dotyczące zjawiska występowania nakładających się genów: ich identyfikacji, skali zjawiska oraz roli i znaczenia tego zjawiska, między innymi w regulacji ekspresji genów.

Jako, że nakładanie się genów jest zjawiskiem bardziej złożone niż na pierwszy rzut oka (intuicyjnie) może się wydawać, dr Rosikiewicz ograniczył swoją analizę (i) do genów nakładających się w orientacji głowa do głowy (końcami 5') oraz (ii) tylko do predefiniowanych genów kodujących białka. Moim zdaniem wprowadzenie takiego ograniczenia było słuszne, a nawet konieczne. Pomogło ono jednoznacznie zdefiniować badane zjawisko i uniknąć zbyt złożonych analiz, które byłyby trudne do śledzenia a ich wyniki byłyby trudne do intuicyjnej interpretacji.

Przykłady występowania nakładających się genów znane były od dawna, jednak traktowane były one jako osobliwe fenomeny, i brakowało odpowiednich narzędzi, które pozwoliłyby na rzetelne oszacowanie ich skali i znaczenia. Ostatnio, pojawiły się takie narzędzia [pełne i dobrze adnotowane sekwencje referencyjne genomów oraz różne aplikacje sekwencjonowania nowej generacji (NGS, ang. *next generation sequencing*)], jednak zjawisko występowania nakładających się genów pozostało poniekąd w cieniu „bardziej modnych” tematów i nie jest mi znana dogłębna analiza tego zagadnienia oraz wyczerpująca lista nakładających się genów.

Dlatego podjęcie się identyfikacji i charakterystyki nakładających się genów u człowieka oraz myszy uważam za jak najbardziej uzasadnione, ważne i niezwykle ciekawe.

Rozprawa ma klasyczną strukturę i składa się z następujących elementów: (i) Streszczeń po polsku i angielsku, (ii) Wstępu, (iii) Celu pracy, (iv) Materiałów, (v) Metod, (vi) Wyników, (vii) Dyskusji, (viii) Wniosków i (ix) Literatury, które są jednocześnie wyróżnionymi rozdziałami rozprawy. Dodatkowo rozprawa uzupełniona jest zestawem tabel dodatkowych, które zawarte są w Aneksie. Jako część rozprawy można traktować również bazę danych OverGeneDB, udostępnioną publicznie w formie portalu internetowego, która powstała i była

wykorzystywana w ramach niniejszej pracy doktorskiej, i jest opisana w ostatnim podrozdziale Wyników.

Ad (i) Jako że streszczenie zostało napisane w mojej ocenie wzorcowo (na jednej stronie w bardzo jasny sposób zostały przedstawione uzasadnienie i cel pracy oraz najważniejsze wyniki) pozwolę sobie pominąć w recenzji streszczania poszczególnych części, a skupię się jedynie na ich ocenie.

Ad (ii) Wstęp bardzo dobrze wprowadza w zagadnienie pracy. W skrótowny lecz wyczerpujący sposób przedstawia zagadnienia, które uzasadniają podjęcie pracy a jednocześnie ułatwiają zrozumienie rozprawy. Co ważne we wstępie brak „rozpisywania się” na tematy tylko luźno powiązane z rozprawą i niewnoszące nic w jej zrozumienie. Przedstawiony wstęp przypomina nieco rozszerzony wstęp dobrej publikacji.

Ad (iii) i (iv) Nie mam uwag do części Materiały i Metody.

Ad (v) Cel pracy, poprzedzony krótkim uzasadnieniem, został sprecyzowany jasno. W zasadniczej części cel pracy ma charakter opisowy (identyfikacja i charakterystyka zjawiska nakładających się genów) i nie polega na testowaniu określonej hipotezy. Nie mniej jednak, jak pisałem wyżej, cała praca jest jak najbardziej uzasadniona i została prawidłowo zrealizowana (odpowiada zakresowi wyników przedstawionych w następnych częściach pracy).

Ad (v), (vi) i (vii) Kolejno przeprowadzone analizy i ich rezultaty przedstawione są w kolejnych podrozdziałach, a ich układ wynika z logicznej konsekwencji i chronologii. Ułatwia to stopniowe zapoznanie się z treścią rozprawy. Poszczególne analizy przeprowadzone są prawidłowo a przyjmowane założenia czy ograniczenia są wyjaśnione i przekonująco uzasadnione. Interpretacja wyników jest prawidłowa, nie dopatrzylem się żadnej nadinterpretacji. Wszystkie analizy wsparte są odpowiednimi testami statystycznymi i generalne wnioski, takie jak np. wyższy poziom ekspresji genów nakładających się, potwierdzone są odpowiednim poziomem istotności statystycznej. Wnioski wynikające z analiz, które nie pozwalają na uogólnienie, takie jak np. wpływ nakładania się genów na alternatywny splicing czy regulację ekspresji genów po transfekcji, potwierdzone są i dobrze zilustrowane bardzo przekonującymi przykładami.

Na wyróżnienie tutaj zasługuje również prezentacja wyników w formie bardzo starannie przygotowanych rycin. Wiele rycin zawiera wyszukane typy wykresów, które dobrze ilustrują przedstawiane zjawiska. Zwraca uwagę ich estetyka, która przejawia się chociażby konsekwencją stosowania kroju i rozmiaru czcionek, grubości linii czy kolorystyki. Raz przyjęty schemat kolorystyczny czy np. schemat rysowania genów/transkryptów stosowane są konsekwentnie przez całą pracę. Niewątpliwie ilustracje zawarte w rozprawie nie są tylko kolorowym przerywnikiem ale znacząco ułatwiają śledzenie i zrozumienie przedstawianych treści. Uwagi te dotyczą również rycin w innych częściach rozprawy.

Do najważniejszych wyników niniejszej rozprawy należą:

- 1) Identyfikacja 582 ludzkich i 113 mysich par nakładających się genów.
- 2) Obserwacja, że nakładanie się genów nie jest cechą stałą, w niewielkim stopniu powtarzalna jest pomiędzy różnymi tkankami, jak również w niewielkim stopniu jest konserwatywna pomiędzy człowiekiem i myszą, nawet w tym samym typie tkanek.
- 3) Obserwacja, że w promotorach genów nakładających się występuje około 2x więcej miejsc wiązania się czynników transkrypcyjnych niż w promotorach pozostałych genów, i identyfikacja czynników transkrypcyjnych, których miejsca wiązania znacząco są nadreprezentowane w promotorach genów nakładających się.

- 4) Obserwacja, że zjawisko nakładania się genów związane jest z wyższą ekspresją genów (nawet w przypadku porównania tych samych genów, które nakładają się w jednych tkankach (próbkach) a nienakładają się w innych).
- 5) Identyfikacja przykładów, w których zachodzenie się genów odgrywa rolę w regulacji ekspresji i alternatywnego splicingu, czy związku zachodzenia się genów z różnymi wzorami modyfikacji histonów oraz wzoru aktywności polimerazy II.
- 6) Przygotowanie bazy danych OverGeneDB, która może posłużyć jako źródło danych do dalszych analiz zjawiska nakładania się genów, jak i informacji na temat wzoru transkrypcji dowolnego genu będącego przedmiotem naszego zainteresowania.

Praktycznie wszystkie wyniki zostały powtórzone zarówno u człowieka jak i myszy co dodatkowo potwierdza ich wiarygodność.

Uwagi i pytania:

W rozdziale 5.2.2. mgr Rosikiewicz zastanawia się nad czynnikami, które mogą wpływać nad zróżnicowaniem liczby nakładających się genów między analizowanymi tkankami/próbkami. Wśród zidentyfikowanych czynników zwiększających liczbę nakładających się genów są: (i) liczba genów ulegająca ekspresji w danej próbce oraz (ii) transfekcja, w przypadku linii komórkowych. Ten pierwszy czynnik, moim zdaniem trochę niesłusznie, mgr Rosikiewicz marginalizuje. Pozostałą zmienność liczby nakładających się genów mgr Rosikiewicz przypisuje w większości specyficzności tkankowej. Chociaż nie można odmówić pewnej słuszności temu ostatniemu założeniu, zastanawiam się, czy nie należałoby uwzględnić również czynników zmienności eksperymentalnej czy biologicznej, które z punktu widzenia prowadzonej analizy, w znacznym stopniu są niekontrolowalną, zmiennością losową. Zresztą czynnik zmienności eksperymentalnej został wymieniony w tytule rozdziału 5.2.2., jednak potem został pominięty, przynajmniej w odniesieniu do ludzkich tkanek. Za udziałem czynnika zmienności eksperymentalnego i/lub biologicznej może świadczyć fakt znacznej różnicy liczby nakładających się genów w dwóch próbkach pochodzących z ludzkiego serca. Również w pewnym stopniu o udziale zmienności eksperymentalnej może świadczyć korelacja liczby nakładających się genów z liczbą genów ulegającej ekspresji w danej próbce, która tylko częściowo może być wytłumaczona tkankowo specyficznym globalnym poziomem ekspresji. Czy na obserwowaną liczbę nakładających się genów nie może wpływać (kontrolowalna lub niekontrolowalna) zmienność takich elementów procedury eksperymentalnej jak: izolacji DNA, generowanie cDNA, generowanie biblioteki, sekwencjonowanie czy liczba i jakość odczytów sekwencjonowania?

W pracy niekonsekwentnie wyrażany jest współczynnik OR (ang. *overlap ratio*), czasami w formie procentów, czasami jako ułamek. Może to wprowadzać w błąd.

Na rycinie 20 rozkład współczynnika JoinedOR, który z definicji powinien zawierać się w przedziale 0-1, wychodzi poza ten przedział. Nie wiem czy dobrze rozumiem ten wykres.

Powyższe uwagi albo (pierwsza uwaga) dotyczą doprecyzowania i mają charakter dyskusji z mojej strony albo (dwie ostatnie uwagi) są bardzo drobne.

W wielu analizach genetycznych, w tym w klinicznej interpretacji mutacji zidentyfikowanych w genach związanych z chorobami człowieka, rutynowo korzysta się z map genów wynikających z sekwencji referencyjnej genomu człowieka i adnotowanych na niej „konsensusowych” transkryptach, najczęściej zdefiniowanych w bazie danych RefSeq. Większość genetyków, w tym ja, patrząc na jakiś gen nie zastanawia się specjalnie nad możliwością występowania innych niż „konsensusowe” miejsc startu transkrypcji, a tym

samym możliwości występowania „nie-konsensusowych” 5'UTR i „nie-konsensusowych” eksonów. W związku z powyższym mam pytanie, do którego chciałbym żeby mgr Rosikiewicz odniósł się w trakcie obrony. Jaka część identyfikowanych, przy okazji wykonywania pracy doktorskiej, miejsc startu transkrypcji nie pokrywała się z miejscami „konsensusowymi”?

Podsumowując, chciałbym zwrócić uwagę, że chociaż zasadnicza część rezultatów, niniejszej pracy doktorskiej nie została jeszcze opublikowana w recenzowanych czasopismach naukowych, to jednak forma ich prezentacji przypomina artykuł naukowy co sugeruje, że wyniki te zostały przygotowane pod kątem ich publikacji, i w mojej ocenie odpowiadają publikacji w prestiżowym czasopiśmie naukowym. Również należy zwrócić uwagę, że zagadnienia poruszane we wstępie do niniejszej pracy zostały również opublikowane w artykule przeglądowym, którego mgr Rosikiewicz jest pierwszym i korespondencyjnym autorem. Ponadto mgr Rosikiewicz jest współautorem 4 innych artykułów opublikowanych w różnych uznanych czasopismach naukowych.

W związku z powyżej przytoczonymi argumentami jestem przekonany, że złożona rozprawa spełnia wszystkie warunki rozprawy doktorskiej i zwracam się do Rady Naukowej Wydziału Biologii UAM w Poznaniu o dopuszczenie mgr Wojciecha Rosikiewicza do dalszych etapów przewodu doktorskiego. Jednocześnie zwracam się do Rady Naukowej Wydziału Biologii UAM o wyróżnienie rozprawy doktorskiej mgr Rosiewicza stosowną nagrodą.



Piotr Kozłowski