

Magda Konarska
czł. koresp. PAN, dr hab. prof. ucz.
ReMedy
Laboratorium Biologii RNA
ul. S. Banacha 2c, 02-097 Warszawa

Warszawa, 11 marca 2021

Ocena osiągnięci naukowego, osiągnięć naukowo-badawczych i aktywności naukowej oraz dorobku dydaktycznego i popularyzatorskiego
doktor Barbary Uszczyńskiej-Ratajczak
pt. „Genomiczna charakterystyka długich niekodujących RNA w genomach człowieka i myszy”
w związku z postępowaniem w sprawie nadania stopnia doktora habilitowanego w dziedzinie nauk ścisłych i przyrodniczych w dyscyplinie nauki biologiczne.

Ocena formalna

Jako osiągnięcie naukowe został zgłoszony cykl trzech powiązanych tematycznie publikacji składających się na dokonanie naukowe pt. „Genomiczna charakterystyka długich niekodujących RNA w genomach człowieka i myszy”. Materiały, otrzymane w języku polskim i angielskim, zawierają: 1) autoreferat; 2) informację o przebiegu kariery, dorobku naukowym, dydaktycznym, wdrożeniowym i popularyzatorskim; 3) wykaz opublikowanych prac i kopie najważniejszych publikacji opisujących osiągnięcia naukowe oraz oświadczenia współautorów wyrażających zgodę na wykorzystanie wieloautorskich prac w postępowaniu o nadanie dr Barbarze Uszczyńskiej-Ratajczak stopnia doktora habilitowanego.

Przedstawione materiały w pełni spełniają wymogi formalne określone w Ustawie z dnia 20 lipca 2018 r „Prawo o szkolnictwie wyższym i nauce”. Wiodący współautor publikacji jednoznacznie określa istotny udział dr Uszczyńskiej-Ratajczak w koncepcji prac. W dwóch pracach eksperymentalnych zgłoszonych jako osiągnięcie naukowe, dr Uszczyńska-Ratajczak jest współ-pierwszym autorem. Ponadto, jest ona pierwszym autorem pracy przeglądowej. Zgłoszone do osiągnięcia prace zostały opublikowane w międzynarodowych pismach o bardzo wysokiej renomie i poziomie oddziaływania: Nature Communications, Nature Genetics, i Nature Reviews – Genetics.

Ocena kariery zawodowej

Kariera naukowa dr Uszczyńskiej-Ratajczak przebiegała kolejno w wielu różnych instytucjach, takich jak: Politechnika Wrocławska, Uniwersytet Adama Mickiewicza w Poznaniu, Instytut Chemii Bioorganicznej PAN, Politechnika Warszawska. Dr Uszczyńska-

Ratajczak spędziła też kilka lat (2013-16) w Centre for Genomic Regulation, CRG, w Barcelonie, w jednym z najlepszych w Europie ośrodków w dziedzinie genomiki, a następnie pracowała w Centrum Nowych Technologii Uniwersytetu Warszawskiego i w Międzynarodowym Instytucie Biologii Molekularnej i Komórkowej w Warszawie. Jej bogata historia działalności w różnych instytucjach świadczy o dynamicznym, zaangażowanym podejściu do pracy naukowej i dobrze wróży na dalsze etapy rozwoju jej kariery naukowej.

Dr Uszczyńska-Ratajczak ukończyła dwa kierunki studiów magisterskich: Biotechnologię na Politechnice Wrocławskiej, Wydział Chemiczny (2008), za pracę pt. „siRNA-mediated gene silencing of integrin beta3 expression inhibits metastatic potential of A459 cells”, kierowaną przez dr Annę Nasulewicz-Goldeman, oraz Bioinformatykę na Uniwersytecie Adama Mickiewicza w Poznaniu, Wydział Biologii (2011), za pracę pt. „Wirtualne laboratorium genomiczne” pod kierunkiem prof. Marka Figlerowicza. Następnie Dr Uszczyńska-Ratajczak ukończyła doktorat z zakresu biochemii w Instytucie Chemii Bioorganicznej PAN, (2013), z rozprawą pt. „Optymalizacja ścieżek analizy danych niestandardowych uzyskanych przy użyciu mikromacierzy DNA” pod kierunkiem prof. Piotra Kozłowskiego. Dodatkowo, uzyskała Dyplom Wydziału Elektroniki i Technik Informacyjnych Politechniki Warszawskiej (2019) za „Projektowanie systemów informatycznych z bazami danych”.

W wyniku pracy w tylu różnorodnych instytucjach w Polsce i za granicą, dr Uszczyńska-Ratajczak zdobyła bardzo ciekawe doświadczenie naukowe doskonale przygotowujące ją do dalszej samodzielnej pracy. Tematyka jej kolejnych etapów pracy nie jest bezpośrednią kontynuacją jednego tematu; jest raczej zbliżona do wspólnego, krystalizującego się w miarę upływu czasu zainteresowania szeroko pojętą analizą transkryptów RNA, przygotowującego kandydatkę do utworzenia samodzielnego zespołu. Szczególnie istotne jest jej doświadczenie zdobyte w Centre for Genomic Regulation, CRG, w Barcelonie, w jednym z wiodących ośrodków badań genomicznych w Europie. Ten staż postdoktorski zaowocował kilkoma bardzo ważnymi publikacjami, trzy z których stanowią osiągnięcie naukowe przedstawione w niniejszej rozprawie habilitacyjnej.

Dr Uszczyńska-Ratajczak jest współautorem 16 prac badawczych i przeglądowych, z czego 12 prac opublikowała już po uzyskaniu stopnia doktora. Są to w większości prace opublikowane w bardzo dobrych, renomowanych pismach naukowych o znaczącym *impact factor* (w szczególności. *Nature Comm*, *Nature Genet*, *Nature Rev Genet*). Ich tematyka dotyczy analizy informatycznej ekspresji RNA, koncentrując się na analizie ekspresji genów niekodujących - obecnie ilość lncRNA w ludzkim genomie jest szacowana na ok. 15-18 tys. i ta liczba będzie prawdopodobnie wzrastać w miarę dołączania wyników nowych, ciągle postępujących badań.

Dr Uszczyńska-Ratajczak uzyskała niezależne finansowanie swoich badań (granty OPUS 13 i POLONEZ 3), jak również była wykonawcą dwóch grantów ENCODE. Poza tym uzyskała stypendia naukowe – dla wybitnego naukowca, z Ministerstwa Nauki i Szkolnictwa Wyższego (2018-21), Nagrodę Dyrektora Instytutu Chemii Bioorganicznej PAN oraz Stypendium dla doktorantów. Dorobek popularyzatorski oraz informacja o współpracy międzynarodowej dr Uszczyńskiej-Ratajczak obejmuje udział w szeregu konferencjach naukowych, w formie wygłaszania prezentacji, udziału w dyskusjach czy prezentacji plakatów. Większość z tych konferencji (10) to znaczące konferencje międzynarodowe (m.in. w Heidelbergu, Barcelonie, Bernie, Londynie czy Hixton), Dr Uszczyńska-Ratajczak prezentowała też plakaty na konferencjach krajowych - we Wrocławiu i Poznaniu.

Bardzo istotnym dorobkiem dr Uszczyńskiej-Ratajczak jest udział w pracach GENCODE Gene Annotation Project, finansowanym przez National Institutes of Health, USA. Z udziału we wszystkich tych aktywnościach widać wyraźnie, że dr Uszczyńska-Ratajczak była bardzo czynna w kształtowaniu swojej działalności naukowej i nawiązywaniu cennych współprac naukowych.

Dr Uszczyńska-Ratajczak była/jest też odpowiedzialna za opiekę naukową nad doktorantami w charakterze opiekuna naukowego lub promotora pomocniczego (w sumie 4 doktorantów), potwierdzając w ten sposób swoją niezależność w pełnieniu funkcji mentora naukowego. Podsumowując, kariera zawodowa dr Barbary Uszczyńskiej-Ratajczak obejmuje szerokie doświadczenie zebrane w różnych ośrodkach w Polsce i za granicą i dowodzi bardzo wysokiego poziomu przygotowania do pracy naukowej. **Ogólnie oceniam bardzo wysoko jej osiągnięcia na tym etapie rozwoju kariery naukowej.**

Ocena osiągnięcia naukowego stanowiącego cykl publikacji

Niniejszą rozprawę habilitacyjną stanowi cykl prac pt. „Genomiczna charakterystyka długich niekodujących RNA w genomach człowieka i myszy”, składający się z trzech prac (2 prace doświadczalne i jedna przeglądowa) opublikowanych w renomowanych czasopismach międzynarodowych. We wszystkich trzech pracach dr Uszczyńska-Ratajczak jest albo pierwszym autorem (praca przeglądowa w *Nature Reviews Genetics*) lub współ-pierwszym autorem (prace doświadczalne w *Nature Communications* i *Nature Genetics*), potwierdzając, że jej wkład w te prace był wiodący. Również autoreferat stanowi bardzo dobrze przygotowane kompendium wiedzy na omawiany temat.

Osobiście chcę podkreślić, że o wiele wyżej cenię sobie osiągnięcie naukowo-badawcze przedstawione w postaci trzech prac o tak wysokim poziomie naukowym niż osiągnięcia, na które składa się wiele pozycji drobnych, mniej istotnych publikacji – chciałabym, żeby stało się to coraz częstszym przypadkiem wśród ocenianych wniosków o habilitację.

Pierwsza z omawianych prac: **J. Lagarde***, **B. Uszczyńska-Ratajczak***, **et al**, *Extension of human lncRNA transcripts by RACE coupled with long read high-throughput sequencing (RACE-Seq)* **Nat Commun**, 2016, 17(7), 12339, opublikowana jest w bardzo wysoko notowanym czasopiśmie (*Impact Factor 12,45*) i koncentruje się na analizie długich niekodujących RNA w genomach człowieka i myszy, budując mapy genowe tych transkryptów. W ramach swoich badań dr Uszczyńska-Ratajczak opracowała udoskonalone ukierunkowane, wysokoprzepustowe metody analizy oparte na sekwencjonowaniu trzeciej generacji za pomocą długich odczytów. Jest to istotny wkład, pozwalający na wysoką dokładność analizy, przy zminimalizowanym wymaganiu rekonstrukcji transkryptów. Działania te doprowadziły też do podniesienia jakości katalogu ENCODE, poprzez poprawę opisu modeli transkryptów już zapisanych w katalogu, a także przez identyfikację kolejnych transkryptów. Poprawa opisu transkryptów dotyczy również dokładniejszego opisu rejonów promotorowych, analizy struktury transkryptów i dokładniejszej analizy ich możliwości kodowania białek.

Praca pozwoliła też na głębsze porównanie metod RACE i CAGE, prowadząc do ciekawych wniosków. Wydłużenie sekwencji genów potwierdzonych metodą CAGE przy użyciu RACE jest raczej rzadkie - większość z nich ma już dość dobrze określone miejsce inicjacji transkrypcji poprzez selekcję RNA zawierających strukturę 5'-cap. Tym niemniej, RACE-seq pozwoliło na identyfikację wielu nieznanymi poprzednio elementów, w tym 615 nowych miejsc startu transkrypcji. Wyniki wykazują, że dane zawarte w katalogu GENCODE v7 są ciągle niekompletne i nie wystarczają do określenia pozycji końców transkryptów. To stwierdzenie nie jest zaskakujące, skoro przeciętny ludzki transkrypt ma ok. 2,5 tys nt, a średnia długość odczytu platformy 454 to tylko 600 nt. Ponieważ nowsze platformy (PacBio czy Oxford Nanopore) pozwalają na znacznie dłuższe odczyty (do 50-200 tys nt), użycie tych platform zostało przyjęte w następnym etapie badań. Ponadto, czułość wykrycia transkryptu przy użyciu RACE-seq jest stosunkowo niska. W tej sytuacji metoda CSL (Capture Long Sequencing) jest bardziej odpowiednia do badania genomiki lncRNA.

Kolejna publikacja: **J. Lagarde*, B. Uszczyńska-Ratajczak*, et al.**, *High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing*, **Nat Genet**, 49(12),1731-1740, 2017, jest poświęcona analizie lncRNA przy użyciu CLS, pozwalającej na nowe spojrzenie na genomową i transkryptomyczną charakterystykę lncRNA. Metoda ta została opracowana w celu rozwiązania niektórych ograniczeń RACE-seq, a ponadto pozwala ona nie tylko na wzbogacenie wybranych sekwencji RNA, ale jednocześnie na odczyt długich sekwencji (PacBio). W ten sposób przebadane zostały rejony genomu w których podejrzewano istnienie nowych loci lncRNA. Analizą objęte zostały niekodujące regiony w ludzkim (ponad 9 tys regionów, ~15,5 megazasad) i mysim (ponad 6,6 tys regionów, ponad 8 megazasad) genomie, co odpowiada odpowiednio 41% i 36% wszystkich lincRNA w GENCODE. Badania objęły również regiony zawierające UCE (ultra conserved elements), enhancers, oraz przewidywane nowe lncRNA. Biblioteki do sekwencjonowania były wzbogacone w wybrane regiony przez amplifikacje z puli cDNA. Ponadto, w wyniku projektu CLS, ulepszona została ocena i systematyczna analiza porównawcza katalogów lncRNA.

Zastosowanie CLS pozwoliło na ogromne zwiększenie ilości zidentyfikowanych transkryptów (ok. 75% z nich to nowe struktury genów lncRNA); m.in. znaleziono nowe transkrypty w przestrzeni międzygenowej. Wyniki te jasno wykazały, że CLS pozwala na nieporównywalnie głębszą analizę genomu, ale jednocześnie podkreśliły obecność istotnych luk w rozumieniu struktury genomu, otwierających przestrzeń do dalszych udoskonaleń systemu. Jednakże imponująca poprawa skuteczności i dokładności analizy CLS pozwala z dużym optymizmem oczekiwać dalszych postępów w tym zakresie, a i już na obecnym etapie najnowsze wersje GENCODE stanowią bardzo ważny krok w kierunku pełnego opisu i charakterystyki lncRNA. Opracowana metoda minimalizuje udział bezpośredniej weryfikacji katalogu przez kuratorów, ponieważ jakość komputerowej obróbki danych dla GENCODE+ jest już bardzo wysoka. Wysokoprzepustowe bioinformatyczne analizy zastosowano na wszystkich etapach analizy danych, od mapowania odczytów, przez tworzenie modeli transkryptów i ocenę jakości. Ponadto, analiza objęła ocenę kompletności 5' i 3' końców transkryptów przy pomocy FANTOM CAGE, DHS, i ustalenia TTS przez identyfikację sekwencji polyA i sygnałów poliadenylacji. Jest to bardzo istotny nowy element opracowanej analizy – pozwalający na analizę alternatywnych miejsc definiujących końce transkryptu. W wielu przypadkach istnienie tego typu informacji staje się niezwykle cenne dla poprawnej analizy wyników i ustalenia właściwego kierunku dalszych badań.

Zastosowanie metod CPAT (Coding Potential Assessment Tool) i PhyloCSF nie wykazało istotnego potencjału kodowania białka, jednak nie wykluczyło też istnienia niekanonicznych ramek odczytu, wykrywanych w niektórych lncRNA. Ponadto, sekwencje promotorowe lncRNA są często zachowane w ewolucji, sugerując funkcje regulatorowe tych RNA w procesie transkrypcji. Struktury promotorów lncRNA i mRNA są podobne pod względem modyfikacji histonów w tych rejonach, ale promotory lncRNA wykazują podwyższony poziom wiązania represyjnych znaczników chromatyny. Wyraźnie lepszy opis lncRNA badanych w projekcie CLS pozwoliło na weryfikację obrazu tych transkryptów jako grupy: ustalono, że lncRNA są znacznie dłuższe niż początkowo sądzono (mediana ponad 1 tys nt) i zawierają po kilka egzonów (średnio ponad 4).

Podsumowując, obie przedstawione prace eksperymentalne stanowią bardzo istotny wkład w proces ciągłego polepszania opisu transkryptów lncRNA, stanowiąc bardzo ważne osiągnięcie naukowe w dziedzinie genomowej i transkryptomicznej analizy RNA.

Praca: **B. Uszczyńska-Ratajczak, J. Lagarde, et al.**, *Towards a complete map of the human long non-coding RNA transcriptome*, **Nat Rev Genet**, 19(9), 535-548, 2018, podsumowuje poprzednio opisane usprawnienia w analizie informatycznej ekspresji lncRNA.

Opisane są różne podejścia eksperymentalne/informatyczne zastosowane w poprzednich dwóch pracach, służące do poprawy katalogów lncRNA i dokładności opisu transkryptów w poszczególnych loci. Opisane są też nowsze modyfikacje procedury, np. wzbogacanie bibliotek przez firmę Roche, pozwalające na wzbogacenie do 100 megazasad sekwencji. Omówione są też ograniczenia metody CLS, wraz z sugerowanymi możliwościami dalszych usprawnień w przyszłych wdrożeniach – np. wzbogacanie transkryptów posiadających 5'-cap czy zastosowanie znacznie lepszych nowych technologii sekwencjonowania trzeciej generacji, umożliwiających zupełnie nowe rozwiązania niedostępne na etapie realizowania projektu CLS. Ponadto, ciągle ulepszane są platformy stosowane przez firmy do sekwencjonowania o wyższej przepustowości i obniżonym koszcie dla użytkownika. Istotną kwestią jest też omówienie stopnia kompletności aktualnych katalogów lncRNA. Wraz z postępem analizy okazuje się, że odkrywane są tysiące nowych loci lncRNA i znane poprzednio egzony łączone są w pojedyncze, dłuższe transkrypty. Istnieją sugestie, że splicing transkryptów lncRNA może przebiegać trochę inaczej niż dla transkryptów kodujących białka i obejmować znacznie wyższy poziom alternatywnego splicingu w obróbce niekodujących RNA – być może chodzi o różnice w zależności alternatywnego splicingu od transkrypcji? Są to bardzo ciekawe kwestie, wyraźnie dowodzące, że w zakresie analizy transkryptomiki lncRNA pozostaje ciągle jeszcze wiele do zrobienia.

Praca wprowadza szereg nowych pojęć pozwalających na bardziej precyzyjne oszacowanie jakości dostępnych katalogów lncRNA, takich jak kompletność transkryptów (ułamek transkryptów o pełnej długości), kompleksowość (liczba skatalogowanych genów), czy poziom wyczerpania locus (liczba transkryptów w danym locus), i porównuje szereg istniejących katalogów lncRNA. Co ciekawe, nie zauważono korelacji między kompletnością i kompleksowością: katalogi o dużej liczbie genów (np. NONCODE czy MiTranscriptome; ok. 62-45 tys genów) posiadają zaledwie ok. 9-4,5% kompletnych transkryptów. Katalogi BIGTranscriptome czy GENCODE+ mają najwyższą liczbę kompletnych transkryptów (27-24%) uzyskały z kolei niski poziom kompleksowości (ok. 12-13,5%). Metoda CLS, jak również najnowsze ulepszenia technologii sekwencjonowania trzeciej generacji dają szansę katalogowi GENCODE na znaczną poprawę kompletności lncRNA w niedalekiej przyszłości. Wynikiem, który zrobił na mnie największe wrażenie jest stwierdzenie, że nawet dwa najpopularniejsze katalogi, GENCODE i RefSeq, zawierają mniej niż 50% wspólnych genów lncRNA. W sumie oznacza to, że bardzo wiele jest jeszcze do zrobienia w celu uzyskania pełnej informacji o transkryptach lncRNA.

Podsumowując, przedstawione publikacje dobitnie dowodzą, że dr B. Uszczyńska-Ratajczak prowadzi istotną działalność naukową o dużym znaczeniu dla szerokiej grupy badaczy zajmujących się lncRNA. Ponadto, prace te mają też duże znaczenie w szerszym sensie – wypracowując nowe rozwiązania pozwalające na pełniejszy opis i charakterystykę innych grup transkryptów. Prace te powstały w przeciągu zaledwie trzech lat (2016-18), a obejmują wiele zasadniczych zmian usprawniających opis analizy lncRNA. Dotychczasowy szybki rozwój dziedziny informatycznej analizy transkryptów pozwala przypuszczać, że w niedługim czasie opisane metody ulegną dalszym ulepszeniom; jako biolog doświadczalny mogę potwierdzić, że takie ulepszone opisy transkryptomów stanowią ogromną pomoc w planowaniu i interpretacji doświadczeń biologicznych. Niewątpliwie, dalsze prace w tym kierunku pozwalają myśleć z optymizmem o przyszłości. Poprawa możliwości technicznych analizy lncRNA pozwoli na sprawniejsze i tańsze badanie ekspresji tych bardzo ciekawych i ważnych dla funkcjonowania komórki transkryptów.

W sumie stwierdzam, że przedstawione osiągnięcie naukowe, opublikowane w renomowanych czasopismach, stanowi bardzo istotny wkład w nasze podejście do genomycznej analizy transkryptomów lncRNA. Prace dotyczą bardzo ważnego wycinka analizy RNA, charakteryzującego się wielką dynamiką prowadzonych badań. Przedstawiony materiał doskonale dowodzi, że dr Uszczyńska-Ratajczak jest bardzo dobrze przygotowana do prowadzenia dalszej samodzielnej pracy.

Ocena działalności dydaktycznej i organizacyjnej

Dr B. Uszczyńska-Ratajczak prowadzi aktywną działalność w zakresie opieki nad młodymi kadrami naukowymi. Obecnie pełni funkcję promotora pomocniczego wobec doktorantki pracującej pod jej bezpośrednią opieką, poprzednio była opiekunem trzech innych doktorantów. Uczestniczyła w wielu zjazdach krajowych i zagranicznych, nawiązała i nadal podtrzymuje bardzo ważną współpracę naukową z członkami zespołu z CRG w Barcelonie. Działalność ta świadczy o jej rozpoznaniu w środowisku naukowym i uznaniu dla jej warsztatu pracy.

Podsumowanie i wniosek końcowy

Na podstawie oceny przedstawionej dokumentacji szczególnego osiągnięcia naukowego (w znaczeniu artykułu 16.2 ustawy), osiągnięć naukowo-badawczych, dorobku dydaktycznego jak również poziomu współpracy międzynarodowej, stanowiących wspólnie podstawę postępowania o nadanie stopnia naukowego doktora habilitowanego, **bardzo wysoko oceniam dorobek naukowy, dydaktyczny i organizacyjny dr Barbary Uszczyńskiej-Ratajczak**. Cykl publikacji dotyczący genomowej charakterystyki długich niekodujących RNA w genomach człowieka i myszy stanowi nowatorski i bardzo ważny wkład do opracowywania metod analizy transkryptomu. Wysoko oceniam również skuteczność ubiegania się kandydatki o fundusze na badania. Całokształt dorobku oraz przedstawione osiągnięcie dr Barbary Uszczyńskiej-Ratajczak stanowią znaczący wkład w rozwój nauki, w pełni odpowiadający kryteriom stawianym przy ubieganiu się o stopień doktora habilitowanego, zgodnie z wymaganiami Ustawy z dnia 20 lipca 2018 r. „Prawo o szkolnictwie wyższym i nauce”. W związku z tym, **wnioskuję do Rady Naukowej Instytutu Chemii Bioorganicznej PAN o nadanie dr Barbarze Uszczyńskiej-Ratajczak stopnia doktora habilitowanego. Jestem przekonana, że rozpatrywane osiągnięcie zasługuje na wyróżnienie, gdyż obejmuje wybitne prace o bardzo wysokim poziomie naukowym i dużym znaczeniu dla rozwoju dziedziny, opublikowane w międzynarodowych czasopismach o bardzo wysokiej renomie i poziomie oddziaływania.**

Z wyrazami szacunku,


Magda Konarska