



UNIwersytet  
Warszawski

Wydział Biologii  
Instytut Genetyki i Biotechnologii  
prof. dr hab. Paweł Golik



Warszawa, 17.10.2024

Recenzja rozprawy doktorskiej mgr Jana Grzegorza Kosińskiego pt. „Identyfikacja i charakterystyka rho-niezależnych terminatorów transkrypcji u bakterii oraz ich udziału w odpowiedzi na stres antybiotykowy”

Terminacja transkrypcji u bakterii jest podstawą wielu charakterystycznych dla tej grupy organizmów mechanizmów regulacji ekspresji genów. Mimo tego, że ogólna charakterystyka miejsc terminacji, zarówno tzw. rho-niezależnej, jak i zależnej od czynnika rho i innych, została opisana już jakiś czas temu, eksperymentalna identyfikacja miejsc terminacji na skalę transkryptomu wciąż pozostaje wyzwaniem i jest przedmiotem aktualnie prowadzonych badań. RNA bakteryjne charakteryzują się krótkim czasem trwania i ulega szybkiej degradacji. Przy zastosowaniu technik transkryptomiki, zarówno uniwersalnej metody RNA-seq, jak i protokołów wyspecjalizowanych w identyfikacji końców 3' (Term-seq), problemem pozostaje rozróżnienie właściwych miejsc terminacji i produktów postępującej od końca 3' degradacji. Opracowanie takich narzędzi bioinformatycznych stanowiło jeden z głównych wątków ocenianej pracy. Doktorant nie poprzestał na stworzeniu odpowiednich programów, ale z powodzeniem wykorzystał je też do scharakteryzowania terminacji transkrypcji u kilkunastu gatunków bakterii, z czego szczególnie ciekawym poznawczo i istotnym z punktu widzenia potencjalnych aplikacji był projekt stworzenia kompletnego katalogu miejsc terminacji rho-niezależnej u *Staphylococcus aureus*. Bakteria ta jest jednym z najistotniejszych klinicznie patogenów człowieka, a szczególnie dotkliwym w jej przypadku problemem jest oporność na wiele antybiotyków. Doktorant podjął się więc zbadania, jak regulacja genów na poziomie terminacji transkrypcji wpływa na odpowiedź *S. aureus* na stres antybiotykowy. Zadania, których rozwiązania podjął się Doktorant spełniają zatem najważniejsze moim zdaniem kryterium – podejmują się rozwiązania naprawdę istotnego poznawczo i oryginalnego problemu naukowego w dyscyplinie nauki biologiczne.

Praca ma formę rozprawy, w skład której wchodzi trzy anglojęzyczne wieloautorskie publikacje, z których dwie już są opublikowane, a jedną przedstawiono w formie preprintu. Dodatkowo, dotychczas nieopublikowane wyniki przedstawione są w formie tradycyjnych rozdziałów rozprawy. Kwestia udziału Doktoranta w publikacjach wieloautorskich jest jasno przedstawiona i potwierdzona dołączonymi do rozprawy oświadczeniami współauterek i współautorów. Doktorant z powodzeniem

połączył teksty publikacji z fragmentami opisującymi jeszcze nieopublikowane wyniki i rozprawa tworzy spójną i logicznie powiązaną całość. Praca prawidłowo cytuje właściwą literaturę i od strony formalnej spełnia wszystkie wymagania stawiane tego typu tekstom naukowym.

Rozpoczynający pracę wstęp zaczyna się zwięzłym wprowadzeniem w zagadnienia ekspresji genów, z szczególnym naciskiem na transkrypcję u bakterii. Integralną częścią wstępu jest opublikowany artykuł przeglądowy (w którym Doktorant jest równorzędnym wiodącym autorem) poświęcony roli struktur drugorzędowych RNA w ekspresji genów bakterii. Dodatkowo, we wstępie znalazło się wprowadzenie do (jeszcze nieopublikowanej) części poświęconej badaniu terminacji transkrypcji u *S. aureus* i problemom antybiotykooporności i roli regulowanej terminacji transkrypcji w rozwoju lekooporności. Ostatnia sekcja wstępu poświęcona została metodom identyfikacji końców 3' RNA u bakterii i wyzwaniom towarzyszącym ukierunkowanej na identyfikację miejsc terminacji transkrypcji analizie danych transkryptomicznych. Po lekturze tej części pracy nie mam żadnych wątpliwości, że Doktorant wykazuje się bardzo rozległą ogólną wiedzą teoretyczną w dyscyplinie nauki biologiczne, obejmującą tak różnorodne zagadnienia, jak biologia molekularna RNA, mechanizmy działania antybiotyków i nabywania przez bakterie oporności, oraz bioinformatyczna analiza danych transkryptomicznych, i z nawiązką spełnia wymagania stawiane przed kandydatami do uzyskania stopnia doktora.

Część opisująca metody i wykorzystane zestawy danych jest dosyć zwięzła, nie stanowi to jednak wady pracy, gdyż w projektach opierających się głównie na tworzeniu narzędzi bioinformatycznych, stworzony algorytm jest raczej wynikiem pracy, i tam też jest opisywany. Bardzo szczegółowy opis działania algorytmu TERMITE, będącego jednym z głównych osiągnięć pracy, zawarty jest zatem w załączonej jako część wyników publikacji. Podobnie, to w wynikach opisane zostały pozostałe stworzone przez Doktoranta narzędzia obliczeniowe.

Dla osiągnięcia wyznaczonego i jasno przedstawionego w pracy celu, Doktorant zrealizował kilka szczegółowych zadań badawczych. W pierwszej kolejności podjął się stworzenia metody analizy danych uzyskiwanych za pomocą protokołów Term-seq, ukierunkowanych na identyfikację końców 3' RNA ze szczególnym uwzględnieniem miejsc terminacji rho-niezależnych. Doktorant zidentyfikował słabości i luki w dotychczas dostępnych programach i stworzył własny, który łączy statystyczną analizę wiarygodności zidentyfikowanych eksperymentalnie stabilnych końców 3' z przewidywaniem i adnotacją miejsc terminacji rho-niezależnej na podstawie charakterystycznych dla nich elementów sekwencji i struktury drugorzędowej. Cały projekt został opisany w preprintie publikacji udostępnionym na platformie bioRxiv. Warto podkreślić, że poza stworzeniem samego narzędzia, publikacja przedstawia też zastosowanie go do identyfikacji miejsc rho-niezależnej terminacji transkrypcji w kilkunastu genomach bakteryjnych, dla których dostępne są zbiory danych

z odczytami sekwencji z protokołu Term-seq oraz porównawczą analizę terminatorów z różnych gatunków.

W kolejnej części, Doktorant dostrzegł problem związany ze stosunkowo małą dostępnością danych pochodzących z eksperymentów Term-seq. Postanowił zatem opracować narzędzie, które mogłoby dostarczyć informacji o miejscach rho-niezależnej terminacji transkrypcji na podstawie bardzo licznych zbiorów danych standardowych analiz transkryptomocnych metodą RNA-seq. Przy tworzeniu narzędzia termiRNAtor Doktorant posłużył się tu algorytmami uczenia maszynowego (konkretnie – klasyfikacją w oparciu o model lasów losowych) zaimplementowanymi w pakiecie scikit-learn. Ta część pracy nie została jeszcze opublikowana, ale kod źródłowy programu został udostępniony na platformie github. Taki sposób udostępnienia kodu dostarczył też dodatkowej weryfikacji tego, że to Doktorant samodzielnie zrealizował zasadniczą część zadania – wszystkie zarejestrowane zmiany w kodzie (*commit*) pochodzą z konta mgr Kosińskiego. Stworzony program został wytrenowany na danych z gatunków, dla których dostępne są pozwalające na bardziej jednoznaczną identyfikację miejsc rho-niezależnej terminacji transkrypcji odczyty protokołów Term-seq, a następnie przetestowany dla oceny czułości i specyficzności.

W kolejnych etapach pracy, Doktorant stanął przed technicznym problemem związanym z tym, że dostępne narzędzia adnotacji elementów genomów są dostosowane głównie do pracy z danymi pochodzącymi z sekwencjonowania genomowego i nie nadają się do pracy ze zbiorami transkryptomocnymi. W odpowiedzi na ten problem, Doktorant stworzył narzędzie nazwane AGouTI, które zostało szczegółowo opisane w dołączonej do rozprawy publikacji, która ukazała się w renomowanym czasopiśmie *PLOS Computational Biology*. Także w tym przypadku program został udostępniony poprzez github, a Doktorant jest autorem wszystkich zarejestrowanych zmian w kodzie.

Kulminacją pracy jest zastosowanie wszystkich stworzonych przez Doktoranta narzędzi bioinformatycznych do stworzenia całogenomowego atlasu rho-niezależnych terminatorów transkrypcji u patogennych szczepów MRSA gronkowca złocistego. Doktorant scharakteryzował zidentyfikowane terminatory i porównał je z wcześniej opisanymi w publikacji poświęconej algorytmowi TERMITE sekwencjami innych gatunków bakterii. Porównując terminację transkrypcji u bakterii poddanych działaniu antybiotyków z kontrolą, Doktorant uzyskał wyniki sugerujące, że regulacja na poziomie warunkowej terminacji/antyterminacji transkrypcji w obrębie operonu *pyr* stanowi jeden z możliwych mechanizmów odpowiedzi bakterii na stres antybiotykowy. Wyniki te są nie tylko bardzo interesujące poznawczo, ale mogą też stanowić punkt wyjścia dla poszukiwania nowych substancji o działaniu przeciwbakteryjnym.

Podsumowując – Doktorant w trakcie pracy rozwiązał szereg istotnych poznawczo i praktycznie problemów naukowych i bez wątplenia wykazał się umiejętnością prowadzenia samodzielnej pracy

naukowej, z dużą nawiązką spełniając wymagania stawiane przed kandydatami do stopnia naukowego doktora. Zarówno sformułowane pytania naukowe, jak i zaproponowane przez Doktoranta rozwiązania są całkowicie oryginalne i w istotny sposób przyczyniają się do postępu wiedzy. Stworzone w ramach recenzowanej pracy narzędzia bioinformatyczne są ogólnie dostępne i stanowią cenny wkład Doktoranta w repertuar programów do analizy danych transkryptomicznych. Jestem przekonany, że będą one szeroko wykorzystywane przez społeczność badaczy zajmujących się transkryptomyczną analizą RNA bakteryjnych.

Jak zawsze w przypadku nietrywialnych prac naukowych, w trakcie jej lektury nasuwają się pewne pytania i uwagi:

- Z opisu algorytmu termiRNAtor wynika, że do identyfikacji potwierdzonych metodą Term-seq miejsc terminacji rho-niezależnej wykorzystano program TERMiTE, który wykorzystuje zarówno odczyty NGS, jak i elementy struktury drugorzędowej i sekwencji. Natomiast cechami wykorzystywanymi bezpośrednio w trenowaniu klasyfikatora, a następnie w analizie są już (zgodnie z opisem na s. 88) elementy sekwencji (konkretne nukleotydy w otoczeniu analizowanego miejsca) i pokrycie odczytami. Czy Doktorant rozważał też wykorzystanie jako cechy podlegającej trenowaniu i przewidywaniu także potencjalnych elementów struktury drugorzędowej (struktury hairpin i ich energia)?
- Na wykresie 6 (s. 107) podpis do panelu b) sugeruje, że przedstawiona jest tam liczba terminatorów niekanonicznych. Tymczasem wykres przedstawia udział procentowy takich terminatorów. Nie jest jasne, czy chodzi o % wszystkich miejsc terminacji (1128 u *S. aureus*), czy % miejsc rho-niezależnych (838 u *S. aureus*)?
- Na s. 108 Doktorant pisze, że udział terminatorów kanonicznych u *S. aureus* jest „stosunkowo wysoki”. Tymczasem z Wykresu 6b wynika, że wyższy jest on u *B. subtilis*, *Enterococcus faecalis* i *Listeria monocytogenes*, a znacząco niższy głównie u przedstawicieli rodzaju *Streptomyces*. Czy w związku z tym udział terminatorów kanonicznych u *S. aureus* nie powinien raczej być opisany jako „średni”?
- Na Wykresie 6d liczba terminatorów odniesiona jest do liczby genów. Biorąc pod uwagę policistronowy charakter bakteryjnych operonów, czy nie należałoby też przedstawić liczbę terminatorów przypadających na jednostkę transkrypcyjną (operon)?
- Na s. 118 Doktorant stwierdza, że obserwowane punkty terminacji u *S. aureus* znajdują się średnio 2161 nukleotydów od najbliższego genu zlokalizowanego poniżej. Czy dotyczy to wszystkich punktów terminacji, czy tylko terminatorów rho-niezależnych? Biorąc pod uwagę wyraźnie prawoskośny charakter rozkładu na Wykresie 12, użycie średniej jest tu mylące, bardziej informatywną wartością będzie mediana. Niezależnie od

tego, znaczna liczba obserwacji wskazuje na dużą odległość między miejscem terminacji transkrypcji a początkiem kolejnej jednostki transkrypcyjnej (górny kwartył powyżej 1500-2000 bp). Co znajduje się w tych obszarach genomu? Czy dużą rolę odgrywają tu miejsca wczesnej terminacji wewnątrz operonu, istotne dla regulacji przez antyterminację?

- Na s. 123 Doktorant stwierdza, że jego obserwacje wskazują na to, że regulacja biosyntezy puryn jest jednym z mechanizmów indukowanych działaniem antybiotyków. Tymczasem, opisywane obserwacje dotyczą operonu *pyr*, który u znanych mi bakterii związany jest z syntezą pirymidyn, a nie puryn.

W podsumowaniu stwierdzam, że przedstawiona mi do recenzji praca doktorska Pana mgr Jana Kosińskiego pt. „Identyfikacja i charakterystyka rho-niezależnych terminatorów transkrypcji u bakterii oraz ich udziału w odpowiedzi na stres antybiotykowy” stanowi oryginalne rozwiązanie istotnego problemu naukowego i potwierdza ogólną wiedzę teoretyczną Kandydata oraz umiejętność samodzielnego prowadzenia pracy naukowej. Na tej podstawie stwierdzam, że recenzowana rozprawa doktorska mgr Jana Kosińskiego spełnia warunki określone w art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2018 r. poz. 1668 z późn. zm.) i wnioskuję do Rady Naukowej Dyscypliny Nauki Biologiczne Uniwersytetu im. Adama Mickiewicza w Poznaniu do dalszych etapów postępowania ws. nadania stopnia doktora w dziedzinie nauk ścisłych i przyrodniczych w dyscyplinie nauki biologiczne.

Biorąc pod uwagę bardzo wysoki poziom merytoryczny rozprawy, potwierdzony publikacją znacznej części wyników w renomowanych czasopismach, a także duży dorobek naukowy Doktoranta, wnioskuję też o wyróżnienie rozprawy mgr Jana Kosińskiego.

prof. dr hab. Paweł Golik