



## UNIwersytet Warszawski

Instytut Informatyki  
ul. Banacha 2  
02-097 Warsaw  
POLAND

dr hab. Anna Gambin, prof UW  
Phone: +(48 22) 5544 566  
Fax: +(48 22) 5544 400  
e-mail: aniag@mimuw.edu.pl

Warszawa, 6.09.2013

### Recenzja rozprawy doktorskiej

**Tytuł rozprawy:** Selection of miRNA isoform markers differentiating between follicular thyroid cancer and follicular thyroid adenoma from high-throughput sequencing data

**Autor rozprawy:** mgr inż. Tomasz Stokowy

Rozprawa zajmuje się bardzo ważnym w diagnostyce chorób tarczycy rozróżnieniem pomiędzy łagodnymi guzkami pęcherzykowatymi (gruczolaki, *ang. follicular adenoma - FA*), a rakiem pęcherzykowym tarczycy (*ang. follicular thyroid cancer - FTC*). Problem ten został uznany za kluczowy przez Europejskie Stowarzyszenie Chorób Tarczycy, gdyż szacuje się, że błędne diagnozy są odpowiedzialne za znaczny procent niepotrzebnych ingerencji chirurgicznych. Badania przeprowadzone przez Autora rozprawy sugerują, że adekwatnymi markerami dyskryminującymi mogą być izoformy mikro RNA (miRNA). Izoformy miRNA są badane dzięki technologii szybkiego sekwencjonowania (HTS). Głównym rezultatem badawczym jest kompleksowa analiza danych molekularnych: poczynając od analizy sekwencji uzyskanych dzięki HTS, poprzez selekcję dyskryminujących cech do finalnej klasyfikacji pacjentów. Dodatkowo zostały przeprowadzone badania danych mikromacierzowych pokazujących ekspresję miRNA w rozróżnianych tkankach. Autorowi udało się wyselekcjonować zbiór 12 cząsteczek miRNA, które wydają się obiecujące jako potencjalne markery w diagnostyce badanych schorzeń. Najbardziej istotne statystycznie miRNA wykazują dodatkowo mniejszą ekspresję w tkankach rakowych.

Wyniki przedstawione w rozprawie były prezentowane w postaci plakatów oraz wystąpień ustnych na bardzo dobrych konferencjach. Dodatkowo Autor wymienia cztery manuskrypty dotyczące problematyki rozprawy, które zostały zgłoszone do publikacji.

Rozprawa napisana w języku angielskim poprawnie i zwięźle. Ta zwięźłość w moim mniemaniu nieznacznie negatywnie wpływa na jakość rozprawy. Ta uwaga nie umniejsza dokonań Autora, którego udział w interdyscyplinarnym międzynarodowym projekcie

RAU	Biuro Dziekana	
	Wpłynęło dnia	16.09.2013
	Nr	967 / zał.

umożliwił uzyskanie bardzo ciekawych wyników. Pragnę tutaj podkreślić zarówno interdyscyplinarność rozprawy, której tematyka łączy zagadnienia medyczne z nietrywialną analizą bioinformatyczną oraz statystyczną jak i potencjalne zastosowania diagnostyczne osiągniętych wyników.

Rozdział pierwszy rozprawy stanowi wprowadzenie do omawianych zagadnień: przedstawia cechy charakterystyczne badanych schorzeń tarczycy oraz prezentuje stosowane w rozprawie metody analizy danych molekularnych. Następnie zostaje sformułowany główny cel rozprawy jakim jest znalezienie markerów miRNA dyskryminujących tkanki rakowe od łagodnych zmian pęcherzykowych gruczolaka.

Kolejny rozdział opisuje badany w rozprawie materiał (próbki uzyskane technologią HTS i mikromacierzową) oraz zastosowane metody walidacji. Trzeba tu podkreślić, że dzięki współpracy z kilkoma krajowymi i zagranicznymi ośrodkami naukowymi Autor dysponował dość licznym zbiorem danych (co jest trudne do uzyskania zwłaszcza w przypadku dość rzadkich przypadków występowania FTC).

Rozdział czwarty przedstawia metody stosowane przez Autora rozprawy do analizy danych: zarówno znane z literatury przedmiotu jak też oryginalne własne metody. Zaprezentowane jest też podejście do porównania danych mikromacierzowych i HTS oraz baza miRBase, z której pochodzą sekwencje referencyjne. Omówiono używane metody klasyfikacji pod nadzorem i bez nadzoru oraz stosowane metody selekcji cech dyskryminujących. Rozdział zawiera też spis używanych znanych z literatury algorytmów analizujących dane HTS oraz krótki opis narzędzi stworzonych przez Autora na potrzeby tej analizy.

W rozdziale piątym przedstawiono uzyskane rezultaty: analizę danych po kontroli jakości, uzyskane sekwencje miRNA oraz ich potencjał dyskryminacyjny; wyselekcjonowane statystycznie istotne cechy oraz wyniki klasteryzacji i klasyfikacji. Okazało się, że wszystkie uzyskane sekwencje miRNA były już znane w literaturze, a wiele z nich wskazywano jako potencjalne markery dyskryminujące złośliwe i łagodne zmiany nowotworowe tarczycy (Tabela 6). Spośród nich wybrano 7 najbardziej specyficznych (Tabela 7). Co ważniejsze dużo lepszą dyskryminację uzyskano na podstawie danych HTS uwzględniających specyficzne izoformy miRNA. Niezależnie wykonano klasteryzację próbek FTC i FA na podstawie globalnego profilu ekspresji za pomocą analizy składowych głównych (PCA) oraz grupowanie hierarchiczne za pomocą algorytmu Warda. Niestety metody te nie pozwoliły na wyraźną separację grup próbek. Podobne wyniki uzyskano dla metody skalowania wielowymiarowego (MDS) gdzie za odległość przyjęto bliskość próbek policzoną metodą losowych lasów (RF). Niesatysfakcjonujące rezultaty analizy bez nadzoru sugerowały, że w problemie klasyfikacji próbek niezbędne jest znalezienie istotnych biomarkerów. Zbiór biomarkerów zaproponowany na podstawie T-testu oraz drzew klasy-

fikujących pozwolił na dużo lepszą dyskryminację (jakość klasyfikatora została oceniona w weryfikacji krzyżowej typu LOOCV).

Ostatni rozdział podkreśla znaczenie technologii HTS w diagnostyce medycznej oraz przedstawia geny, które mogą być regulowane przez wyselekcjonowane w pracy miRNA.

Rozprawa zrobiła na mnie dobre wrażenie dowodząc opanowania przez Autora warsztatu badawczego w stopniu czyniącym zadość wymaganiom stawianym rozprawom doktorskim. Rozważane w pracy zagadnienie naukowe jest bardzo ważne i zostało wyczerpująco zbadane przez Autora. O szerokiej wiedzy świadczy przeprowadzona analiza literatury przedmiotu. Znaczna część wyników zawartych w rozprawie stanowi oryginalny dorobek Autora.

Drobne uwagi, jakie nasunęły mi się w trakcie czytania pracy dotyczą staranności, czytelności i zbytniej lakoniczności w prezentowaniu stosowanych metod:

- W rozdziale prezentującym metody badawcze, algorytmy i narzędzia będące oryginalnym wkładem Autora powinny zostać wyeksponowane i zaprezentowane oddzielnie od znanych z literatury metod.
- Podpisy pod większością rysunków powinny być bardziej specyficzne i czytelne.
- Rozdzielczość rysunków 36 i 37 nie pozwala odczytać formuł umieszczonych przy osiach Y.

Wymienione powyżej uwagi nie wpływają na ocenę merytorycznych wyników zaprezentowanych w rozprawie, które klasyfikuję wysoko. Podsumowując stwierdzam, że recenzowana przeze mnie praca spełnia wymagania stawiane rozprawom doktorskim przez obowiązujące przepisy i wnoszę do Rady Wydziału Automatyki, Elektroniki i Informatyki Politechniki Śląskiej o dopuszczenie mgr inż. Tomasza Stokowego do dalszych etapów przewodu doktorskiego.



  
dy. hab. Anna Gambin