

## Ocena rozprawy doktorskiej

**Autor rozprawy:** mgr Anna Papież

**Tytuł rozprawy:** Integrative data analysis methods in multi-omics molecular biology studies for disease of affluence biomarker research

**Promotor rozprawy:** prof. dr hab. Inż. Joanna Polańska

### Syntetyczne przedstawienie pracy

Rozprawa doktorska pani Anny Papież przedstawiona jest w postaci angielskojęzycznej monografii obejmującej 122 strony obustronnego maszynopisu. Wyniki przedstawiono na 38 rycinach i w 11 tabelach. Praca podzielona jest na 5 rozdziałów: 1 wstęp, 2 tło (background), 3 materiały i metody, 4 wyniki i dyskusja, 5 wnioski. Struktura jest logiczna i dobrze odzwierciedla treść pracy.

Celem pracy było opracowanie różnych metody ujednoczenia i integracji wyników otrzymanych z badań o wysokiej przepustowości (high throughput analyses) w dziedzinie biologii molekularnej, ze szczególnym naciskiem na biomarkery osobniczej promieniowrażliwości. Autorka opracowała metodę rozpoznania i korekty efektu próbki (batch effect) i zastosowała ją do integracji wyników eksperymentalnych, których celem było znalezienie markerów promieniowrażliwości i ekspozycji na promieniowanie na poziomie ekspresji genów (transkryptomika), poziomu białek (proteomika) i połączonej analizy wyników z obu metod (inter-omics). Eksperymentalna część pracy jest poprzedzona przeglądem badań i problemów w dziedzinie biomarkerów i integracji wyników badań o wysokiej przepustowości. Pracę zamyka dyskusja i osobny rozdział poświęcony wnioskowi.

### Opinia o poprawności i oryginalności postawionej tezy i w jakim stopniu została ona wykazana

Tezą pracy jest założenie konieczności opracowania metod do integracji wyników badań o wysokiej przepustowości. Teza ta jest absolutnie poprawna w sytuacji, gdzie badania o wysokiej przepustowości są coraz częściej stosowane. Równocześnie, problem efektu próbki jest często pomijany zarówno przez autorów badań jak i recenzentów manuskryptów, co prowadzi do publikacji wyników o ograniczonej wartości poznawczej. Mgr Papież jasno opisuje problem i uzasadnia tezę. Oryginalnym aspektem jest opracowanie algorytmu opartego na dynamicznym programowaniu.

### Opinia o analizie źródeł świadczącej o dostatecznej wiedzy autora w danej dyscyplinie naukowej

Wiedzę autorki oceniam jak bardzo dobrą; rozprawa obejmuje 146 pozycji literaturowych w skład których wchodzi prace dotyczące wszystkich poruszonych aspektów. Na określenie zasługują też dobrze napisane rozdziały wstępne, opisujące zagadnienia poruszone w rozprawie.

### Opinia o pozycji rozprawy w stosunku do stanu wiedzy reprezentowanych przez literaturę światową

Badania o wysokiej przepustowości są szeroko prowadzone a analiza ich wyników wymaga specjalnego warsztatu statystycznego. Wkład autorki do opracowania nowych metod analizy statystycznej jest oryginalny i ważny, czego udowadniają 4 publikacji naukowe, na których opiera się rozprawa doktorska (wyszczególnione na stronie 119 rozprawy).

### Opinia o znaczeniu uzyskanych wyników dla danej dyscypliny naukowej

Eliminacja efektu próbki stanowi niezwykle ważne i trudne wyzwanie w badaniach o wysokiej przepustowości, szczególnie w dziedzinie medycyny precyzyjnej, która opiera się na identyfikacji i

Biuro Dziekana	
Wpłynęło dnia	08.03.2019
nr	393 / zakt.

walidacji biomarkerów. W dziedzinie badań radiacyjnych i ochrony radiologicznej poszukiwane są biomarkery promieniowrażliwości i ekspozycji na promieniowanie. Wyniki mgr Papież mają ważne znaczenie dla dyscyplin medycyna i ochrona zdrowia.

### **Opinia o umiejętności autora do poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników**

Umiejętność tę oceniam wysoko. Autorka jasno opisała metodykę badań, zarówno tę dotyczącą części statystycznej jak i biologicznej. Dyskusja wyników jest dobra i wyczerpująca. Zdolność jasnego opisu interdyscyplinarnych elementów pracy (obliczeniowego i biologicznego) zasługuje na podkreślenie.

### **Główne wady rozprawy**

W przedstawionej pracy autorka skupia się na opisie opracowania i walidacji metod do integracji wyników badań o wysokiej przepustowości. To stanowi zasadniczy element rozprawy autorki która jest biostatystykiem i uważam, że z tego zadania autorka wywiązała się doskonale. Moje zastrzeżenia dotyczą aspektu aplikacji opracowanej metodyki i uwagi krytyczne skupią się na tym problemie, który leży w poszerzonym obszarze problematyki znanej w informatyce jako „GiGo” (Garbage in, Garbage out). **Równocześnie pragnę bardzo wyraźnie podkreślić, że uwagi te dotyczą spraw wychodzącym poza ramy rozprawy doktorskiej i w żadnym stopniu nie osłabiają jej bardzo dobrej oceny.** Chęć ten problem jednak naświetlić, ponieważ autorka porusza się w obszarze badań interdyscyplinarnych i opracowana przez nią metoda analizy danych ma tylko sens we współpracy z przedstawicielami badań biologiczno-medycznych, którzy dostarczają dane. Wydaje mi się, że ocena rozprawy doktorskiej jest dobrym momentem do poruszenia tej kwestii, ponieważ autorka znajduje się na początku ścieżki naukowo-badawczej i ważnym jest, żeby jej dorosłe życie naukowe cechowało krytyczne podejście do otrzymywanych danych wejściowych.

Biomarkery, na których skupia się autorka dotyczą chorób i cech określanych jako złożone lub wieloczynnikowe. Przykładem choroby wieloczynnikowej jest cukrzyca (zarówno typu I jak i II) a przykładem cechy wieloczynnikowej jest osobnicza promieniowrażliwość. Zarówno cukrzyca jak i wysoka promieniowrażliwość mogą też mieć podłoże monogenowe, gdzie cecha kliniczna jest wynikiem mutacji w pojedynczym genie. Takie przypadki są jednak rzadkie a diagnoza cechy klinicznej jest prosta i nie wymaga szukania biomarkerów. U ogromnej większości osób cukrzyca i fenotyp promieniowrażliwości ma jednak charakter złożony, gdzie przyczyny cechy klinicznej wynikają z 1) interakcji wielu alleli o niskiej penetrancji; 2) czynników środowiskowych i 3) przypadku (polecam tu pracę Burga i Lehnera: beyond genotype to phenotype: why the phenotype of an individual cannot always be predicted from their genome sequence and the environment that they experience. FEBS J. 279 (20):3765-3775, 2012). Złożoność cechy wieloczynnikowej powoduje, że znalezienie biomarkera jest niezwykle trudne i wymaga zbierania próbek od dużej liczby pacjentów lub dawców. Badania o wysokiej przepustowości dostarczają ogromnej liczby wyników a opracowana przez autorkę metoda rozwiązania problemu próbki działa potęgująco. W efekcie, stosunkowo łatwo jest uzyskać wynik znamieny dla danego zbioru próbek, który jednak okazuje się niepowtarzalny przy analizie innego zbioru próbek. Przykładem problemu są badania GWAS (genome wide association study). Opracowanie metody GWAS rozbudziło ogromne nadzieje na uzyskanie biomarkerów podatności na choroby wieloczynnikowe, które jednak okazały się płonne (vide praca Burga i Lehnera). W przypadku poszukiwań czynników genetycznych warunkujących osobniczą wrażliwość na promieniowanie dochodzi problem stochastycznego rozkładu „letalnych trafień” promieniowania, na skutek których komórki i tkanki obumierają. Ważne jest, żeby autorka zdała sobie sprawę z dużej liczby czynników wpływających na wystąpienie niepożądanych efektów radioterapii, które nie mają nic wspólnego z podłożem genetycznym pacjenta, na którym skupiły się prowadzone przez nią badania. Podam konkretny przykład: w akapicie 3.2.1 (strona 46 rozprawy) podano odsyłacz do pracy Yarnolda i wsp., jako źródła sposobu identyfikacji pacjentów promieniowrażliwych. Praca Yarnolda i wsp. dotyczy jednak analizy wystąpienia efektów ubocznych w wyniku eskalacji dawki frakcyjnej. U pacjentów napromienionych wyższą dawką frakcyjną oczywiście występuje podwyższone ryzyko odczynów, co

jednak nie ma nic wspólnego z osobniczą promieniowrażliwością. Dalej, na rycinie 3.1 pokazano schemat klasyfikacji pacjentów na „promieniowrażliwych” i „promienioopornych”. Rozumiem, że pacjenci promieniowrażliwi to ci, u których wystąpiły efekty uboczne. Wystąpienie efektów ubocznych po radioterapii w żadnym stopniu nie jest prostą oznaką wysokiej promieniowrażliwości. Na prawdopodobieństwo ich wystąpienia wpływają dawka (patrz praca Yarnolda i wsp.), czynniki genetyczne, środowiskowe (life style factors), przypadek i stochastyczny rozkład trafień letalnych. Ogromna liczba uzyskanych profili ekspresji genów pozwala znaleźć różnice między dwoma grupami pacjentów, które mogą być błędnie interpretowane jako potencjalne biomarkery osobniczej promieniowrażliwości. Na przestrzeni ostatnich 30 lat powstały dziesiątki prac opisujących potencjalne biomarkery osobniczej promieniowrażliwości. Próby ich walidacji na innych grupach pacjentów kończyły się jednak niepowodzeniem (AGIR. Human Radiosensitivity. Report of the Independent Advisory Group on Ionising Radiation, Chilton: Doc HPA, RCE-21, 2013; Rajaraman, M. Hauptmann, S. Bouffler, and A. Wojcik. Human individual radiation sensitivity and prospects for prediction. Annals of the ICRP doi.org/10.1177/0146645318764091, 2018).

Na rycinie 4.10c autorka pokazuje obiecujący biomarker na poziomie „inter-omics” ze 100% swoistością i czułością. Rozumiem, że wynik taki uzyskano dzięki obróbce danych dla danej grupy dawców. Ważne jest, żeby autorka zdawała sobie sprawę, że dla cech złożonych nie istnieją predykcyjne biomarkery o tak wysokiej swoistości i czułości (L. Jostins and J. C. Barrett. Genetic risk prediction in complex disease. Hum.Mol.Genet. 20 (R2):R182-R188, 2011). Z tego powodu pojęcie „personalised medicine” zastąpiono pojęciem „precision medicine” ([http://www.nap.edu/catalog.php?record\\_id=13284](http://www.nap.edu/catalog.php?record_id=13284)). Chodzi w tym o to, że pojęcie „personalised medicine” może przez pacjenta zostać błędnie zrozumiane jako podejście dopasowane konkretnie do niego i dające pewność wyleczenia. Tymczasem jest to podejście, które dla pojedynczego pacjenta oznacza większe niż przeciętne prawdopodobieństwo sukcesu terapii, które jednak nigdy nie wynosi 100%. Inną ważną rzeczą jest, że korelacja między cechą pacjenta a poziomem znalezionej biomarkera nie oznacza od razu, że marker ten posiada wartość predykcyjną.

Powtarzam, że uwagi te dotyczą spraw wychodzącym poza ramy rozprawy doktorskiej, ale szkoda że nie zostały (choćby na marginesie) poruszone w rozprawie doktorskiej.

### **Krytyczne uwagi szczegółowe**

Podczas czytania tekstu znalazłem kilka drobnych błędów, które wyszczególniam poniżej. Błędy te nie wpływają na bardzo dobrą ocenę pracy.

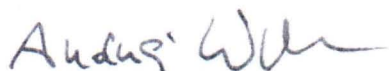
1. Na stronie 17 (wiersz 13) autorka używa pojęcia „personalized medicine” a na stronie 35 (wiersz 20) pojęcia „precision medicine”. Uwaga: pojęcia te mają różne znaczenia ([http://www.nap.edu/catalog.php?record\\_id=13284](http://www.nap.edu/catalog.php?record_id=13284)) i lepiej nie używać ich wymiennie.
2. Strona 21, pierwsze zdanie. Autorka poprawnie definiuje pojęcie „biomarker” jako indyktor zmian chorobowych. Jednak już w 4 zdaniu autorka błędnie nazywa biomarkerem test na krew, mocz, limfocyty i szklisko zęba. Test nie jest biomarkerem, tylko badaniem.
3. Strona 23, wiersze 8-11. Autorka sugeruje (nie podając źródła), że promieniowanie jonizujące jest czynnikiem silnie wywołującym choroby nowotworowe i choroby układu krążenia. To stwierdzenie jest błędne. Wyśredniony dla wieku i płci nadmiar ryzyka względnego dla dawki 1 Gy na całe ciało wynosi dla zgonów na choroby nowotworowe ok. 0,5 a dla chorób układu krążenia ok. 0,14 (K. Ozasa, Y. Shimizu, A. Suyama, F. Kasagi, M. Soda, E. J. Grant, R. Sakata, H. Sugiyama, and K. Kodama. Studies of the mortality of atomic bomb survivors, Report 14, 1950-2003: an overview of cancer and noncancer diseases. Radiat.Res. 177 (3):229-243, 2012). W przeliczeniu na ryzyko bezwzględne, oznacza to ok. 10% dla nowotworów i ok. 4% dla chorób układu krążenia dla dawki 1 Gy. Wartości te spadają o 50% po zastosowaniu czynnika DDREF rekomendowanego przez ICRP dla narażenia chronicznego (ICRP publikacja 103). Mało kto otrzymuje dawkę 1 Gy na całe ciało, również w medycynie, gdzie najczęstszą przyczyną narażenia są badania diagnostyczne. Mamy tu do czynienia z dawkami w zakresie mGy. Ryzyko bezwzględne dla nowotworów i chorób układu

krażenia po dawce 10 mGy podanej w krótkim czasie to, odpowiednio, ok 0,1 i 0,04%. Problem narażenia na promieniowanie jonizujące i ocena ryzyka jest bardzo złożony, dlatego należy wystrzegać się prostych stwierdzeń jak to zrobiono na stronie 23, wiersze 8-11. Jako wstęp do problematyki polecam: <http://www.unscear.org/unscear/en/publications/booklet.html>.

4. Strona 23, wiersze 21-22. Autorka podaje przykłady efektów ubocznych radioterapii nie powołując się na żadne źródła. Lista jest błędna, ponieważ łączy efekty narażenia całego ciała na wysokie dawki (biegunka i wymioty) z efektami chemioterapii o niskim znaczeniu dla jakości życia (utrata włosów) i efektami nie związanymi z terapią (przeżuty i zapalenia stawów). Ostrożnie z prostymi stwierdzeniami, niepopartymi literaturą źródłową!
5. Strona 36, wiersz 14. Zdanie „P-value combination is a statistical concept that was first introduced by Fisher (2006)” sugeruje, że koncepcja ta została wprowadzona w 2006 roku. Czy nie wcześniej?

### **Wniosek**

Na podstawie powyższej recenzji wnioskuję do Wysokiej Rady Wydziału Automatyki, Elektroniki i Informatyki Politechniki Śląskiej **o dopuszczenie rozprawy doktorskiej** do publicznej obrony. Równocześnie **wnioskuję o wyróżnienie rozprawy** ze względu na opublikowanie jej elementów w 4 publikacjach naukowych w czasopismach międzynarodowych, z których, w 3 przypadkach, kandydatka jest pierwszym autorem.



Prof. dr hab. Andrzej Wójcik