

Oczapowskiego 1A, 10-719 Olsztyn-Kortowo, Polska, Telefon: +48 89 523 32 01

dr hab. Agata Żmijewska, prof. UWM  
Katedra Anatomii i Fizjologii Zwierząt  
Wydział Biologii i Biotechnologii  
Uniwersytet Warmińsko-Mazurski w Olsztynie  
10-719 Olsztyn  
Ul. Oczapowskiego 1A  
agata.zmijewska@uwm.edu.pl

Olsztyn, dn. 11 sierpnia 2026 r.

**Ocena rozprawy doktorskiej lek. wet. Klaudii Miętkiewskiej pt. „Transcriptome analysis of early stages of bovine embryos obtained *in vivo* from Polish Holstein-Friesian cows” wykonanej pod kierunkiem prof. dr hab. n. wet. Chandry S. Pareek**

Przedstawiona do oceny rozprawa doktorska prezentuje wyniki badań dotyczących wpływu wieku dawczyń oraz sezonowych zmian środowiskowych na profil transkryptomiczny wczesnych zarodków bydła pozyskanych *in vivo* od krów rasy polskiej holsztyńsko-fryzyjskiej. Istotnym elementem pracy była identyfikacja genów o zróżnicowanej ekspresji oraz procesów biologicznych związanych z rozwojem zarodków z wykorzystaniem nowoczesnych metod sekwencjonowania RNA i zaawansowanych analiz bioinformatycznych, w tym analizy dekonwolucyjnej. Pomimo intensywnych badań prowadzonych w ostatnich latach, wiedza dotycząca molekularnych mechanizmów odpowiedzialnych za wpływ wieku matki oraz warunków środowiskowych na kompetencję rozwojową zarodków bydła nadal pozostaje niepełna. Podjęta problematyka ma istotną wartość poznawczą i potencjalne znaczenie dla rozwoju badań nad biologicznymi uwarunkowaniami jakości zarodków, doskonalenia metod wspomaganego rozrodu oraz poprawy efektywności hodowli bydła.

**Ocena formalna**

Przedstawiona do oceny rozprawa doktorska jest opracowaniem w języku angielskim, które obejmuje 188 stron maszynopisu, w tym 29 rycin oraz 1 fotografię, 16 tabel, i 320 pozycji piśmiennictwa anglojęzycznego. Praca zawiera podstawowe elementy wymagane dla rozprawy doktorskiej, tj. spis treści, abstrakt w języku angielskim, wprowadzenie, hipotezy badawcze, materiał i metody, wyniki oraz ich omówienie w dyskusji, podsumowanie, oraz piśmiennictwo, a także materiały uzupełniające do rozprawy (w postaci ośmiu plików tabelarycznych typu Excel i pięciu plików zawierających ryciny). W rozprawie doktorskiej zabrakło streszczenia w języku polskim, mimo że zgodnie z ustawą, do rozprawy doktorskiej przygotowanej w języku obcym należy dołączyć streszczenie w języku polskim. W mojej ocenie tytuł rozprawy jest zbyt ogólny i nie odzwierciedla w pełni zakresu przeprowadzonych badań, w szczególności dwóch głównych czynników eksperymentalnych, tj. wieku dawczyń

oraz warunków sezonowych, analizowanych w kontekście ich wpływu na profil transkryptomyczny zarodków *in vivo*.

### **Ocena merytoryczna**

**Streszczenie** zostało przygotowane w sposób przejrzysty i obejmuje wszystkie podstawowe elementy rozprawy, tj. cel badań, zastosowane metody, wyniki oraz podsumowanie. W sposób logiczny przedstawiono zakres przeprowadzonych analiz transkryptomicznych oraz analiz bioinformatycznych. W mojej ocenie streszczenie jest jednak nadmiernie rozbudowane poprzez szczegółowe wyliczenie dużej liczby genów o zróżnicowanej ekspresji, co utrudnia wyeksponowanie najważniejszych osiągnięć pracy oraz głównych wniosków wynikających z przeprowadzonych badań. Zdecydowanie bardziej zasadne byłoby skoncentrowanie się na najistotniejszych wynikach oraz ich znaczeniu biologicznym, pozostawiając szczegółowe zestawienia genów do omówienia w rozdziale *Wyniki*.

**Wstęp** (7 stron) został przygotowany w sposób logiczny i przejrzysty. Podzielono go na cztery podrozdziały obejmujące najważniejsze zagadnienia niezbędne do zrozumienia tematyki rozprawy. Doktorantka przedstawiła aktualny stan wiedzy dotyczący wpływu wieku dawczyń oraz stresu cieplnego na rozwój zarodków bydła, omówiła możliwości wykorzystania technologii RNA-seq oraz zagadnienia związane z analizą transkryptomu. Pozytywnie należy ocenić również próbę uzasadnienia wyboru analizowanych czynników (wieku i sezonu) oraz przedstawienie znaczenia badań zarówno dla rozrodu bydła, jak i rozwoju metod wspomaganego rozrodu. Pewne zastrzeżenia budzi jednak sposób przedstawienia analizy dekonwolucyjnej. Doktorantka początkowo opisuje klasyczną dekonwolucję stosowaną do oszacowania udziału różnych typów komórek w próbkach po bulk RNA-seq, podczas gdy w niniejszej pracy zastosowano jedynie podejście inspirowane dekonwolucją do oceny podobieństwa profili transkryptomicznych pomiędzy kolejnymi etapami rozwoju zarodków. Sposób i cel zastosowanego podejścia zostały wyjaśnione dopiero w końcowej części podrozdziału. W mojej ocenie, aby uniknąć mylnego wrażenia dotyczącego zastosowanej metody, taka informacja powinna zostać jednoznacznie przedstawiona już na początku tego rozdziału.

**Hipotezy badawcze** pracy są zgodne z zakresem tematycznym rozprawy, jednak hipoteza pierwsza została sformułowana nieco zbyt szeroko w odniesieniu do zastosowanych metod badawczych. Przeprowadzone analizy transkryptomiczne pozwalają wykazać różnice w ekspresji genów oraz wskazać potencjalnie zmienione procesy biologiczne, natomiast nie umożliwiają bezpośredniej oceny mechanizmów fizjologicznych i metabolicznych. W mojej ocenie hipotezę 1 należałoby sformułować z większym naciskiem na analizę zmian transkryptomicznych. **Cele pracy** odpowiadają tematyce rozprawy i zakresowi przeprowadzonych badań, jednak sposób ich przedstawienia w mojej ocenie mógłby być bardziej dopracowany. Ich opisy są nadmiernie rozbudowane i zawierają liczne powtórzenia dotyczące analizy transkryptomu, zastosowania technologii bulk RNA-seq oraz identyfikacji genów o zróżnicowanej ekspresji. Ponadto część celów szczegółowych (cele 1 i 2) ma charakter przede wszystkim techniczny i opisuje kolejne etapy analizy bioinformatycznej, a nie cele naukowe. Dodatkowo, moim zdaniem, cel 5 warto byłoby doprecyzować, podkreślając, że przeprowadzone analizy pozwalają na ocenę zależności pomiędzy zmianami profilu transkryptomicznego a etapem rozwoju zarodków, a nie na bezpośrednią ocenę wpływu zmian ekspresji genów na rozwój zarodków. Takie uszczegółowienie lepiej odzwierciedlałoby zakres uzyskanych wyników.

**Material i metody** zostały przedstawione w 15-stronicowym rozdziale, zawierającym pięć podrozdziałów obejmujących charakterystykę materiału biologicznego i projekt doświadczenia, procedurę pobierania zarodków i ich oceny, procedury laboratoryjne związane z izolacją RNA i przygotowaniem bibliotek do sekwencjonowania oraz analizę bioinformatyczną danych RNA-seq oraz analizę dekonwolucyjną profili transkryptomycznych. Pozytywnie należy ocenić część metodyczną dotyczącą doboru zwierząt, pozyskiwania zarodków oraz ich oceny zgodnie z kryteriami IETS. Czytelność rozdziału zwiększają również tabele 1 i 2, przedstawiające charakterystykę materiału badawczego oraz podział zarodków na analizowane grupy, a także rycina 1 z reprezentatywnymi obrazami zarodków oraz rycina 2 ilustrująca schemat doświadczenia i przebieg analizy RNA-seq. Jednocześnie zwracam uwagę, że dopiero na schemacie eksperymentu pojawiają się określenia „1st Lactation”, „Beyond VI Lactation”, „Thermoneutral Condition” oraz „Elevated Temperature”, które nie zostały wcześniej wprowadzone ani wyjaśnione w tekście metodyki. W mojej ocenie należałoby ujednolicić nazewnictwo lub jednoznacznie wskazać, że oznaczenia te odpowiadają grupom opisanym w części „Materiały i metody”, co zwiększyłoby spójność opracowania i ułatwiło interpretację schematu. W czasie obrony prosiłabym także o doprecyzowanie pewnych kwestii przedstawionych poniżej. Czy liczebność grup doświadczalnych, zwłaszcza grupy młodszych krów, obejmującej jedynie dwa osobniki mogła wpłynąć na możliwość uogólniania uzyskanych wyników dotyczących wieku dawczyń? W Tabeli 3 przedstawiono charakterystykę jedynie 12 zarodków, podczas gdy zgodnie z wcześniejszym opisem do analiz wykorzystano 34 zarodki. Warto wyjaśnić, według jakiego kryterium wybrano zarodki zaprezentowane w tabeli umieszczonej w rozprawie. Ponadto nie jest jednoznacznie określone, w którym dniu po inseminacji wykonywano płukanie macicy i pozyskiwano zarodki. Biorąc pod uwagę, że inseminację przeprowadzano trzykrotnie w odstępach 12-godzinnych, warto doprecyzować, od którego zabiegu liczono wiek zarodków oraz w jaki sposób uwzględniono to podczas oceny ich stadium rozwojowego. Ze względu na fakt, że w wielu miejscach rozprawy obserwowane różnice sezonowe są interpretowane w kontekście podwyższonej temperatury i stresu cieplnego, zasadne byłoby doprecyzowanie, czy poza pomiarem temperatury otoczenia oceniano również wskaźniki pozwalające potwierdzić występowanie stresu cieplnego u dawczyń, takie jak Temperature-Humidity Index (THI) lub parametry fizjologiczne zwierząt (np. temperatura ciała). Informacje te pozwoliłyby lepiej uzasadnić interpretację obserwowanych różnic sezonowych jako efektu obciążenia cieplnego u samic dawczyń, a nie wyłącznie efektu sezonowego.

Na szczególną uwagę zasługuje sposób opracowania części dotyczącej analizy bioinformatycznej. Doktorantka przedstawiła kolejne etapy przetwarzania danych RNA-seq od kontroli jakości odczytów, poprzez mapowanie, kwantyfikację ekspresji genów, normalizację danych i analizę różnicowej ekspresji, aż do interpretacji funkcjonalnej uzyskanych wyników. Na pozytywną ocenę zasługuje również wskazanie wykorzystania uznanych i powszechnie stosowanych narzędzi bioinformatycznych, takich jak STAR, RSEM, DESeq2, SVA czy Cytoscape, a także zdeponowanie surowych danych sekwencyjnych w bazie NCBI, co umożliwia niezależną weryfikację badań. Jednakże w rozdziałach metodyki dotyczących analiz bioinformatycznych, Doktorantka również nie ustrzegła się nieścisłości. Wyjaśnienia wymaga m.in. opis genomu referencyjnego wykorzystanego do mapowania odczytów. W metodyce wskazano „Bos taurus UMD3.0 (assembly accession GCA\_002263795.3)”, jednak numer akcesyjny GCA\_002263795.3 odpowiada genomowi referencyjnemu ARS-UCD1.3. Warto zweryfikować nazwę zastosowanego genomu referencyjnego oraz zachować spójność pomiędzy jego nazwą a numerem akcesyjnym. W podrozdziałach 3.4.3 oraz 3.5.1 częściowo powielono opis wstępnej kontroli jakości danych sekwencyjnych (primary QC), obejmującej ocenę jakości odczytów oraz kontrolę danych po ich oczyszczeniu. Powoduje to pewne dublowanie informacji i nieco zaburza przejrzystość opisu kolejnych etapów analizy.

W mojej ocenie pełny opis kontroli jakości danych FASTQ warto byłoby zamieścić w jednym miejscu, natomiast rozdział poświęcony analizie bioinformatycznej rozpocząć od właściwego przetwarzania i analizy danych po zakończeniu etapu kontroli jakości. Warto również doprecyzować informacje dotyczące bibliotek określonych jako posiadające „atypical QC patterns” lub „atypical sequencing yield”, wskazując, czy takie przypadki rzeczywiście wystąpiły oraz jaki był ich wpływ na dalszy przebieg analizy. Zastrzeżenia budzi także Tabela 4, ponieważ zgodnie z definicją parametr „Effective (%)” oznacza stosunek liczby odczytów po oczyszczeniu do liczby odczytów surowych, jednak dla wszystkich bibliotek przyjmuje wartość 100,00%, mimo zastosowania procedur usuwania adapterów i sekwencji o niskiej jakości. Dodatkowego wyjaśnienia wymagają również biblioteki nr 13 i 15, które zawierają odpowiednio około 5,6 (raw data 6.8) oraz ~7,9 mln odczytów (raw data 10.2), podczas gdy pozostałe biblioteki obejmują najczęściej od 50 do ponad 80 mln odczytów. Tak znaczna różnica w głębokości sekwencjonowania wymaga komentarza. W opisie analizy bioinformatycznej zabrakło numerów wersji większości wykorzystanych programów (FastQC, fastp, STAR, RSEM, MultiQC, SAMStat, RNA-SeQC, tximeta, DESeq2, SVA). Informacje te są istotne z punktu widzenia odtwarzalności analiz. Doktorantka kilkakrotnie wskazuje również, że zastosowano „recommended/default parameters”, jednak nie precyzuje, których programów dotyczy to stwierdzenie. Ponieważ w analizie wykorzystano wiele narzędzi bioinformatycznych, warto byłoby jednoznacznie określić zastosowane parametry lub odwołać się do konkretnych wersji oprogramowania. W rozdziale *Material i metody* oraz w streszczeniu Doktorantka deklaruje wykorzystanie baz Panther i STRING do analizy funkcjonalnej genów o zróżnicowanej ekspresji, natomiast w rozdziale *Wyniki* nie przedstawiono jednak rezultatów uzyskanych z wykorzystaniem tych narzędzi. Zauważyłam również inną niespójność między rozdziałami *Material i metody* oraz *Wyniki*. W metodyce wskazano, że wyniki analiz zostały zaimportowane do środowiska R z wykorzystaniem pakietu *tximeta*, natomiast w kolejnym rozdziale podano, że zestawienia ekspresji genów uzyskano za pomocą pakietu *tximport* („gene-level summaries were derived via *tximport*”). Ponieważ są to dwa różne pakiety, pełniące częściowo odmienne funkcje, warto jednoznacznie wskazać, które narzędzie zostało ostatecznie wykorzystane odpowiednio ujednolicić opis metodyki i wyników.

Pozytywnie należy ocenić próbę zastosowania analizy dekonwolucyjnej do oceny podobieństwa transkryptomów zarodków do referencyjnych etapów rozwoju przedimplantacyjnego oraz ostrożne podkreślenie przez Doktorantkę, że uzyskane wyniki odzwierciedlają podobieństwo profili transkryptomicznych, a nie rzeczywiste stadium rozwojowe zarodków. Pomimo iż Doktorantka starała się przygotować tę część rozdziału w sposób uporządkowany, w mojej ocenie wymaga ona kilku istotnych uzupełnień i doprecyzowań. Przede wszystkim wyjaśnienia wymaga wybór ludzkiego zbioru referencyjnego (GSE36552) do analizy transkryptomów zarodków bydła. Doktorantka nie uzasadniła, dlaczego zdecydowała się na wykorzystanie danych pochodzących od człowieka. W metodyce podano jedynie, że zachowano geny posiadające jednoznaczne ortologi, jednak nigdzie nie przedstawiono liczby genów wykorzystanych ostatecznie w analizie, kryteriów ich wyboru, ani odsetka genów utraconych podczas mapowania ortologów oraz zastosowanych parametrów pakietu MuSiC, co utrudnia ocenę przeprowadzonej dekonwolucji. Ponadto pomiędzy rozdziałami *Material i metody* oraz *Wyniki* występuje inna niespójność. W metodyce opisano analizę dekonwolucyjną z wykorzystaniem algorytmów MuSiC i DeTREM, natomiast w rozdziale *Wyniki* przedstawiono również analizy ekspresji genów markerowych, hierarchiczne grupowanie próbek, analizę UMAP oraz diagram Venna genów markerowych. W metodyce brakuje opisu sposobu wykonania tych analiz, zastosowanych narzędzi bioinformatycznych oraz kryteriów identyfikacji genów markerowych, przez co uniemożliwia to odtworzenie całego pipeline'u bioinformatycznego w przypadku dekonwolucji. Wątpliwości budzi również opis analizy czułości z wykorzystaniem algorytmu DeTREM. W metodyce

wskazano, że zgodność wyników MuSiC i DeTREM posłużyła do oceny stabilności uzyskanych oszacowań, jednak w rozdziale *Wyniki* zgodność ta została opisana jedynie jakościowo. W konsekwencji analiza dekonwolucji powinna być traktowana jako eksploracyjna ocena podobieństwa transkryptomicznego, a nie jako niezależna metoda klasyfikacji stadium rozwojowego analizowanych zarodków.

Rozdział *Wyniki* stanowi najobszerniejszą część rozprawy (57 stron) i obejmuje rezultaty uzyskane z wykorzystaniem szerokiego spektrum analiz bioinformatycznych. Na pozytywną ocenę zasługuje bogata dokumentacja graficzna wyników, obejmująca wizualizacje różnicowej ekspresji genów oraz analizy funkcjonalne. Doktorantka podjęła próbę kompleksowego opracowania uzyskanych danych transkryptomicznych, prezentując wyniki na wielu poziomach, od kontroli jakości danych, poprzez identyfikację genów różnicowo ekspresjonowanych, aż do ich interpretacji biologicznej. Świadczy to o nakładzie pracy włożonym zarówno w przygotowanie analiz bioinformatycznych, jak i opracowanie uzyskanych rezultatów.

Jednocześnie w mojej ocenie rozdział ten mógłby zostać uporządkowany zarówno pod względem struktury, jak i sposobu prezentacji wyników. Opis wyników zawiera elementy metodyczne (np. opis mapowania odczytów z wykorzystaniem STAR, użyciem RSEM, pakietu *tximport*, parametry analizy ClueGO - wartość współczynnika kappa, itd.), które lepiej byłoby umieścić w rozdziale *Materiał i metody*. Rozdział *Wyniki* zyskałby na przejrzystości, gdyby skoncentrowano się w nim przede wszystkim na prezentacji uzyskanych rezultatów, natomiast szczegóły dotyczące opisu zastosowanych narzędzi bioinformatycznych pozostawiono w części metodycznej.

Moje wątpliwości wzbudził opis wyników analizy genów o zróżnicowanej ekspresji oraz analizy funkcjonalnej (rozdziały 4.2 oraz 4.3). W metodyce wskazano, że geny o zróżnicowanej ekspresji (DEG) identyfikowano na podstawie skorygowanych wartości  $p$  ( $FDR \leq 0,05$ ), natomiast w rozdziale *Wyniki* przedstawiono liczbę genów spełniających kolejne progi nominalnych wartości  $p$  oraz odniesiono się do progu  $P \leq 0,05$ . Dobrze byłoby jednoznacznie wskazać, czy prezentowane wyniki odnoszą się do wartości nominalnych, czy FDR, oraz podać liczbę genów spełniających przyjęte kryterium istotności. Również informacja, że liczba genów określonych jako up-regulowane (6023 lub 7167) i down-regulowane (5548 lub 5585) sumuje się dokładnie do liczby wszystkich genów pozostawionych po filtracji, sugeruje, że podział ten odnosi się do wszystkich analizowanych genów, a nie wyłącznie do genów o ekspresji zróżnicowanej statystycznie. Dla zachowania spójności z metodyką warto jednoznacznie określić kryteria kwalifikacji genów jako DEGs (np. FDR i  $\log_2\text{FoldChange}$ ) oraz zweryfikować zgodność opisu z metodyką i materiałami uzupełniającymi.

W mojej ocenie najważniejsza krytyczna uwaga dotyczy rozdziałów 4.2.1 oraz 4.2.2. Analizując wyniki przedstawione w Tabelach 5 i 6 oraz porównując je z materiałami uzupełniającymi, stwierdziłam istotne nieścisłości i błędy. Część genów przedstawionych w Tabelach 5 i 6 jako istotnie zróżnicowana nie została odnaleziona w przesłanych materiałach uzupełniających lub nie spełnia tam kryteriów istotności statystycznej (np. *ICOS*, *CYP11A1*, *FAM13C*, *PDE10A*, *NPSR1*, *SYT9*, *ARL13A*, *RBMS3* i in.). W Tabelach 5 i 6 w porównaniu z materiałami uzupełniającymi nie zgadzają się również wartości takich parametrów jak *transcript count*, *log2Fold change*, *pvalue* lub *padj*. Jest to szczególnie istotne, ponieważ geny wytypowane w Tabelach 5 i 6 stanowią podstawę dalszej interpretacji biologicznej oraz są szczegółowo omawiane w rozdziale *Dyskusja*. Nie można wykluczyć, że rozbieżności te wynikają z omyłkowego dołączenia niewłaściwej wersji materiałów uzupełniających, ale w obecnej postaci uniemożliwiają pełną weryfikację przedstawionych wyników.

Ponadto chciałabym przedstawić kilka uwag i sugestii odnoszących się do wyników umieszczonych w podrozdziałach 4.2.3 oraz 4.3.3. W jednym podrozdziale przedstawiono łącznie wykresy służące kontroli jakości danych (density plot, PCA), diagnostyki modelu statystycznego (histogram wartości  $p$ , Q-Q plot) oraz prezentacji wyników analizy różnicowej (volcano plot i MA plot), mimo że pełnią one odmienne funkcje. W mojej ocenie rozdział zyskałby na przejrzystości, gdyby poszczególne typy analiz zostały wyraźniej rozdzielone i omówione zgodnie z ich funkcją w procesie analizy danych.

W trakcie czytania rozprawy zwracają także uwagę liczne nieścisłości dotyczące opisu i interpretacji poszczególnych rycin (rozdziały 4.2.3 i 4.3.3), których przykłady podaję poniżej. W przypadku ryciny 3 w opisie umieszczonym w rozprawie wskazano, że przedstawiono znormalizowane dane po transformacji logarytmicznej, podczas gdy podpis ryciny odnosi się do surowych odczytów przed normalizacją. Doktorantka napisała, że po kontroli jakości do dalszych analiz zakwalifikowano 34 biblioteki odpowiadające 34 transkryptomom zarodków, natomiast na rycinie 4 (wykresie PCA) przedstawiono jedynie 32 próbki. Jeżeli dwie próbki zostały wykluczone z analizy, informacja ta powinna zostać jednoznacznie przedstawiona w metodyce lub wynikach wraz z uzasadnieniem ich wykluczenia. Dodatkowo opis ryciny 4 wymaga doprecyzowania. Stwierdzenie o „clear separation” pomiędzy grupami wydaje się zbyt kategoryczne i nie wynika jednoznacznie z przedstawionego wykresu PCA. Bardziej właściwe byłoby użycie określeń sugerujących częściową separację próbek. Nie wyjaśniono również znaczenia kolorów oznaczających Group 1 i Group 2 ani sposobu wyznaczenia tych grup. Z podpisu ryciny wynika jednocześnie, że próbki klasyfikowano według stadium rozwojowego (różne symbole), podczas gdy analiza dotyczyła wieku dawczyń, co może utrudniać interpretację biologicznego znaczenia obserwowanego rozdziału prób.

W doświadczeniu analizowano 34 próbki, dlatego standardowa analiza, przedstawiona na rycinie 5 powinna generować maksymalnie 33 składowe główne, jednak na rycinie widoczne są 32 składowe. Prosiłabym o wyjaśnienie sposobu przeprowadzenia analizy (np. pominięcie składowych zerowych) lub zweryfikowanie poprawności przygotowanej ryciny. Dodatkowo opis rycin 5 i 16 w rozprawie jest niespójny z przedstawionymi wykresami. Zarówno w tekście, jak i w podpisach rycin podano, że PC1, PC2 i PC3 wyjaśniają odpowiednio 29,7%, 13,7% i 8,9% całkowitej wariancji. Tymczasem na obu wykresach wartości wynoszą około 22,0%, 11,7% i 7,2%. Prosiłabym o weryfikację, które wartości są prawidłowe. Ponadto ryciny 5 i 16 wydają się identyczne, co sugeruje, że wykorzystano ten sam wykres zarówno dla analizy wieku, jak i sezonu. Jeżeli było to zamierzone, powinno zostać jednoznacznie wyjaśnione w rozprawie.

Opisy rycin 6–8 również wymagają doprecyzowania. Histogram rozkładu wartości  $p$  (Rycina 6) może wskazywać na obecność sygnału różnicowej ekspresji, jednak sam w sobie nie stanowi dowodu na wysoką jakość ani wiarygodność przeprowadzonej analizy. W przypadku Ryciny 7 nie wyjaśniono, jakie dane wykorzystano do konstrukcji wykresu Q-Q (np. wartości  $p$  czy inne parametry), co utrudnia ocenę zasadności jego zastosowania. Z kolei volcano plot (Rycina 8) identyfikuje geny charakteryzujące się największymi zmianami ekspresji i istotnością statystyczną, jednak nie pozwala samodzielnie wnioskować o ich znaczeniu biologicznym, które powinno wynikać z dalszych analiz funkcjonalnych. Ponadto zapis „after filtering ( $P < 1.0$ )” na str. 66 jest metodologicznie niepoprawny, ponieważ filtracja ekspresji nie była wykonywana na podstawie wartości  $p$ .

Kolejne zastrzeżenia dotyczą Ryciny 9 (MA plot). Kryteria wyboru genów oznaczonych na wykresie nie zostały przedstawione. W tekście wyników wskazano, że zaprezentowano geny o największych zmianach ekspresji („top 25 upregulated genes ranked by  $\log_2$  fold change”), a w podpisie pod ryciną „top 20 differentially expressed genes”, jednak po porównaniu ryciny 9 z Tabelami 5 i 6 stwierdzono, że na wykresie oznaczono jedynie część genów należących do tej grupy, pomijając między innymi *ICOS*, *CYP11A1*, które należą do genów o najwyższych

wartościach bezwzględnego  $\log_2\text{FoldChange}$ . W związku z tym zarówno opis ryciny, jak i tekst wyników nie są zbieżne i wymagają doprecyzowania poprzez jednoznaczne określenie kryteriów wyboru genów przedstawionych na wykresie.

Rozdział *Wyniki* zawiera również liczne powtórzenia. Najważniejsze geny zostały opisane w podrozdziałach dotyczących Top 30 genów, następnie ponownie omówione w rozdziałach „Transcriptomic signatures...”, a ich funkcje przedstawiono jeszcze raz w tabelach 7–10 oraz 13–16. Powoduje to częściowe powtarzanie tych samych informacji oraz zaciera granicę pomiędzy rozdziałami *Wyniki* i *Dyskusja*. Rozbudowana interpretacja funkcji biologicznych oraz wnioski dotyczące potencjalnego wpływu zmian ekspresji na kompetencję zarodków, implantację czy rozwój embrionalny mają charakter dyskusji i powinny zostać przeniesione lub istotnie skrócone.

Wartościowymi elementami rozdziału *Wyniki* są podrozdziały 4.2.6 i 4.3.6, poświęcone analizie funkcjonalnej genów o zróżnicowanej ekspresji. Przeprowadzona przez Doktorantkę analiza umożliwiła identyfikację procesów biologicznych potencjalnie związanych z obserwowanymi zmianami transkryptomycznymi, stanowiąc istotne uzupełnienie analizy różnicowej ekspresji genów. Jednocześnie kilka elementów tej części rozprawy wymaga doprecyzowania, co zwiększyłoby jej przejrzystość oraz ułatwiłoby interpretację przedstawionych wyników. Przede wszystkim warto zachować większą spójność pomiędzy rozdziałami *Materiał i metody* oraz *Wyniki*. Jak wspomniałam wcześniej, w części metodycznej podano, że analizy funkcjonalne przeprowadzono z wykorzystaniem narzędzi Panther, STRING oraz ClueGO, natomiast w rozdziale *Wyniki* przedstawiono wyłącznie wyniki uzyskane z wykorzystaniem ClueGO. Warto wyjaśnić, czy analizy wykonane w Panther i STRING nie dostarczyły istotnych wyników i dlaczego nie przedstawiono ich w wynikach.

W kilku miejscach pojawiają się również nieścisłości dotyczące interpretacji analiz GO i KEGG. Doktorantka wielokrotnie określa przedstawione wyniki jako „GO Biological Process and KEGG pathway enrichment”, jednak zarówno podpisy rycin 10–13 oraz 21–24 oraz przedstawione opisy w rozprawie, zgodnie z moją wiedzą, odnoszą się wyłącznie do terminów Gene Ontology Biological Process. Jeżeli analiza obejmowała również ścieżki KEGG, powinny one zostać jednoznacznie oznaczone na rycinach i w opisie lub wyraźnie oddzielone od terminów GO. Ponadto należy wyjaśnić znaczenie symbolu „\*” na rycinach 11, 13, 22 oraz 23. Zgodnie z metodologią (str. 29) wszystkie terminy przedstawione w analizie ClueGO spełniają kryterium istotności statystycznej po korekcji wielokrotnych porównań (adjusted  $p < 0,05$ ), dlatego nie jest jasne, dlaczego tylko wybrane kategorie oznaczono gwiazdką. Niektóre sformułowania stosowane w opisie wyników w rozdziałach 4.2.6.2 oraz 4.2.6.4, również wymagają doprecyzowania. Dla przykładu, na str. 51 nie jest jasne, co Doktorantka rozumie pod pojęciem „11 metabolic pathways”. Analiza materiałów uzupełniających wskazuje, że przedstawiono serię funkcjonalnych sieci ClueGO dla wybranych genów, dla których kryterium wyboru również nie zostało opisane, obejmujących procesy rozwojowe, immunologiczne, sygnalizacyjne oraz metaboliczne. Dlatego określenie „11 metabolic pathways” wydaje się zbyt wąskie i błędne (ponieważ w Supplementary Figure 4 wskazano 12 sieci). Bardziej adekwatnym określeniem byłoby np. „ClueGO functional networks” lub „pathway enrichment networks”. Dodatkowo niektóre opisy rycin (szczególnie 12, 21–23) zawierają informacje, które nie są w pełni zgodne z prezentowanymi wynikami lub sprawiają wrażenie przeniesionych z innych części analizy, dlatego należałoby je zweryfikować i ujednolicić z przedstawionymi wykresami.

Na pozytywną ocenę zasługuje podjęcie przez Doktorantkę próby zastosowania analizy dekonwolucji do oceny heterogeniczności transkryptomicznej zarodków. Wykorzystanie nowoczesnych narzędzi bioinformatycznych oraz zewnętrznego zbioru referencyjnego świadczy o ambitnym podejściu do analizy danych. Jednocześnie kilka elementów tej części

wymaga doprecyzowania. Rycina 25 w przejrzysty sposób ilustruje zróżnicowanie poziomu ekspresji wybranych genów pomiędzy poszczególnymi stadiami rozwoju zarodkowego, jednak w opisie brakuje uzasadnienia wyboru analizowanych genów oraz liczby próbek (n) dla poszczególnych stadiów rozwoju. W mojej ocenie interpretacja wyników zaprezentowanych na rycinach 26–28 również wymaga większej ostrożności. W przypadku ryciny 26 przedstawione dane nie wskazują jednoznacznego rozdziału wczesnych i późnych stadiów rozwoju, ponieważ stadium 4-komórkowe nie grupuje się z pozostałymi wczesnymi etapami. Ponadto nie wyjaśniono, jakie dane wykorzystano do konstrukcji dendrogramów. Również interpretacja UMAP (rycina 27), metody, która nie została opisana w metodyce, wydaje się zbyt kategoryczna, gdyż służy ona przede wszystkim do wizualizacji podobieństwa próbek i nie stanowi formalnej walidacji. Warto także wyjaśnić, dlaczego na rycinie 28 uwzględniono jedynie grupę 2-/4-komórkową, podczas gdy w poprzednich analizach ta wczesna grupa obejmowała również oocyt i zygotę. Prosiłabym również o wyjaśnienie, do jakich wartości odnosi się procentowy udział markerów przedstawiony na powyższym wykresie, co nie zostało opisane w rozdziale *Wyniki*.

Interpretacja wyników dekonwolucji przedstawionych na rycinach 29 i 30 również budzi wątpliwość. We wszystkich próbkach dominuje sygnatura późnej blastocysty, natomiast sygnatury stadium 8-komórkowego i moruli są praktycznie niewidoczne lub widoczne w niewielkim stopniu. W związku z tym nie jest jasne, w jaki sposób analiza dekonwolucji potwierdza wspomnianą we wstępie i metodyce heterogeniczność rozwojową zarodków oraz jaki był jej wpływ na dalszą interpretację wyników dotyczących wieku dawczyń i sezonu. Warto byłoby uzupełnić tę część o ilościowe zestawienie wyników (brak danych numerycznych) oraz krótki komentarz wyjaśniający biologiczne znaczenie uzyskanych rezultatów w kontekście wpływu wieku dawczyń lub sezonu na profil transkryptomiczny zarodków.

Na zakończenie omawiania rozdziału *Wyniki* chciałabym zaznaczyć, że materiały uzupełniające wymagają ponownej, kompleksowej weryfikacji pod względem spójności, kompletności oraz zgodności z treścią rozprawy. Stwierdziłam szereg nieścisłości dotyczących zarówno numeracji i opisu materiałów uzupełniających z tekstem rozprawy, jak i ich zawartości. Dotyczy to między innymi nieciągłej numeracji próbek w Supplementary Table S1, niejednoznacznego opisu danych przedstawionych w Supplementary Table S2 (FPKM/TPM - przedstawiono tylko jedną macierz ekspresji, przez co nie jest jasne, czy są to wartości FPKM, TPM czy inny sposób normalizacji danych), rozbieżności pomiędzy liczbą próbek deklarowaną w tekście a liczbą kolumn w tabeli, błędnych odwołań do tabel i rycin uzupełniających (np. str. 25, 51, 54, 67, 80), a także niespójności pomiędzy zawartością Supplementary Tables i Figures a ich opisem w rozdziale *Wyniki*. Uchybienia te wpływają nie tylko na przejrzystość i odbiór rozprawy, lecz również utrudniają ocenę poprawności przeprowadzonych analiz oraz weryfikację prezentowanych wyników. W mojej ocenie wszystkie materiały uzupełniające oraz odwołania do nich powinny zostać ponownie zweryfikowane przed przygotowaniem publikacji naukowych, gdyż stanowią integralną część dokumentacji uzyskanych wyników.

Rozdział **Dyskusja** (48 stron, podzielony na 3 główne podrozdziały związane z oceną wpływu wieku dawczyń, sezonu oraz dekonwolucji na profil transkryptomiczny zarodków) wydaje się być przygotowany z dużym zaangażowaniem i świadczy o szerokim przeglądzie literatury wykonanym przez Doktorantkę. Widoczne jest dążenie do możliwie szerokiego omówienia biologicznego znaczenia zidentyfikowanych genów oraz osadzenia uzyskanych wyników w kontekście aktualnej wiedzy. Świadczy to o dużym nakładzie pracy. W wielu miejscach przedstawione rozważania są interesujące i pokazują próbę wyjaśnienia potencjalnego znaczenia genów, dla których dostępne są jedynie ograniczone dane

literaturowe. Doktorantka szczegółowo omówiła każdy z genów o zróżnicowanej ekspresji, jednak większy nacisk należało położyć na próbę wyjaśnienia biologicznego w oparciu o wyniki analiz GO, których praktycznie nie omówiono w tym rozdziale. W efekcie Dyskusja ma momentami charakter przeglądu funkcji poszczególnych genów, a w mniejszym stopniu stanowi krytyczną interpretację własnych wyników Doktorantki. W mojej ocenie takie całościowe spojrzenie pozwoliłoby lepiej powiązać uzyskane wyniki z postawionymi hipotezami badawczymi, wyeksponować oryginalność przeprowadzonych badań oraz podkreślić biologiczne znaczenie uzyskanych rezultatów.

Zwraca również uwagę, że w dyskusji w niewielkim stopniu odniesiono się do wyników analizy dekonwolucji, mimo że stanowią one jeden z elementów przeprowadzonych analiz bioinformatycznych. Uzupełnienie tej części pozwoliłoby lepiej powiązać wszystkie uzyskane wyniki oraz pełniej przedyskutować ich znaczenie w kontekście wpływu wieku dawczyń i warunków sezonowych na profil transkryptomiczny zarodków. Przedstawione wyniki wskazują, że większość próbek została przypisana głównie do sygnatury późnej blastocysty, co nie wydaje się w pełni zgodne z obserwacjami morfologicznymi i danymi umieszczonymi w bazie NCBI PRJNA1026627, a w konsekwencji ogranicza wartość dekonwolucji w ocenie heterogeniczności badanych zarodków. Warto byłoby szerzej omówić ograniczenia wynikające z zastosowania międzygatunkowego zbioru referencyjnego (ludzkie scRNA-seq) oraz zgodność wyników uzyskanych metodami MuSiC i DeTREM.

Rozdział *Conclusions* stanowi podsumowanie najważniejszych wyników uzyskanych w pracy, jednak w mojej ocenie mógłby mieć bardziej syntetyczny charakter. W obecnej formie Doktorantka ponownie omawia zmiany ekspresji poszczególnych genów oraz ich potencjalne znaczenie biologiczne, przez co rozdział przypomina częściowo kompilację wyników i dyskusji. W mojej ocenie większą przejrzystość tej części zapewniłoby skoncentrowanie się na kilku najważniejszych, wypunktowanych wnioskach wynikających bezpośrednio z przeprowadzonych badań oraz wyraźniejsze odniesienie ich do postawionych celów i hipotez badawczych. W związku z tym podczas publicznej obrony prosiłabym Doktorantkę o przedstawienie w sposób syntetyczny najważniejszych wniosków wynikających z przeprowadzonych badań. Przydatnym uzupełnieniem rozprawy byłby również kompleksowy schemat podsumowujący najważniejsze wyniki badań, przedstawiający w sposób syntetyczny wpływ wieku dawczyń oraz warunków sezonowych na profil ekspresji genów, zidentyfikowane procesy biologiczne oraz ich potencjalne znaczenie dla rozwoju zarodków. Taka rycina mogłaby ułatwić zrozumienie wzajemnych zależności pomiędzy uzyskanymi wynikami oraz lepiej wyeksponować najważniejsze osiągnięcia pracy.

Zastrzeżenia budzi również dobór części piśmiennictwa wykorzystanego w rozprawie. Analiza cytowanych publikacji wykazała, że niektóre z nich nie odnoszą się bezpośrednio do omawianego zagadnienia lub nie stanowią właściwego źródła dla przedstawionych informacji. Przykładowo, pozycja 10 dotyczy komórek Sertoliego świni poddanych stresowi cieplnemu, a nie jakości zarodków bydła, pozycje 17 i 43 odnoszą się do raka jelita grubego, pozycja 42 dotyczy morfologii liści pomidora, pozycja 45 opisuje komórki chimer ludzkich i mysich, natomiast pozycja 47 nie zawiera danych dotyczących różnic ekspresji genów pomiędzy zarodkami męskimi i żeńskimi, mimo że została wykorzystana jako źródło takich informacji. W kilku przypadkach odwołania bibliograficzne nie odpowiadają opisywanym narzędziom bioinformatycznym lub zastosowanym metodom. Przykładowo, opis algorytmu MuSiC został oparty na publikacji dotyczącej MuSiC2, która stanowi rozwinięcie oryginalnej metody, natomiast przy opisie programu FastQC przywołano pozycje bibliograficzne dotyczące zagadnień związanych z kontrolą jakości sekwencjonowania, a nie publikację opisującą samo narzędzie. Ponadto dla części wykorzystywanych metod i programów (m.in. *tximeta*, *TMM*,

voom) brakuje odpowiednich cytacji. Również w rozdziale Dyskusja odnaleziono przykłady nieprawidłowego przypisania źródeł literaturowych. Publikację 87 wykorzystano jako źródło informacji o genie SERPINB4, mimo że dotyczy ona chemokiny CCL24, publikacje 100 i 101 nie potwierdzają przedstawionych informacji dotyczących odpowiednio EGR2 i ZBBX, prace 103 i 104 jedynie częściowo uzasadniają wnioski odnoszące się do LRCH2, natomiast do interpretacji wyników dotyczących ARL13A wykorzystano publikacje opisujące ARL13B. Powyższe przykłady wskazują na konieczność ponownej weryfikacji części cytowań oraz ich zgodności z treścią rozprawy, szczególnie w odniesieniu do interpretacji biologicznej uzyskanych wyników.

W rozprawie występują błędy edytorskie i redakcyjne. Obejmują one m.in. błędy stylistyczne, literówki, powtórzenia fragmentów tekstu, niekonsekwentnie stosowaną terminologię (np. „bulk scRNA-seq” na str. 33, 54 i in.), a także pojedyncze nieścisłości w podpisach rycin oraz opisach tabel, co obniża przejrzystość rozprawy i utrudnia weryfikację przedstawionych wyników.

Dodatkowe pytania i sugestie, które nasunęły się w trakcie czytania niniejszej rozprawy doktorskiej.

1. Które spośród zidentyfikowanych genów wybrałaby Pani do dalszych badań funkcjonalnych?
2. Jakie podejście zastosowałaby Pani do niezależnej walidacji wyników RNA-seq i które geny wybrałaby Pani do takiej walidacji? Proszę uzasadnić kryteria ich wyboru.
3. W jaki sposób wyniki analizy transkryptomicznej (zidentyfikowano więcej zmian związanych z warunkami sezonowymi niż z wiekiem dawczyń) można byłoby powiązać z oceną kompetencji rozwojowej zarodków? Jakie eksperymenty zaproponowałaby Pani ?
4. Jakie są najważniejsze ograniczenia zastosowania ludzkiego zbioru scRNA-seq jako danych referencyjnych do analizy transkryptomów zarodków bydła i w jaki sposób można byłoby potwierdzić (zwalidować) uzyskane wyniki dekonwolucji.

### **Wniosek końcowy**

Na podstawie przedstawionej rozprawy można stwierdzić, że Doktorantka posiada wiedzę, umiejętności oraz kompetencje niezbędne do prowadzenia badań naukowych w zakresie biologii molekularnej i analiz transkryptomicznych. Na szczególne podkreślenie zasługuje szeroki zakres przeprowadzonych badań, obejmujący wykorzystanie modelu *in vivo*, analizę RNA-seq oraz zastosowanie nowoczesnych narzędzi bioinformatycznych do analizy dużych zbiorów danych. Przedstawione wyniki mają charakter oryginalny i dotyczą wpływu wieku dawczyń oraz warunków sezonowych na profil transkryptomiczny zarodków. Szczególnie interesujące są wyniki dotyczące wpływu warunków sezonowych na transkryptom zarodków bydła, które mogą stanowić podstawę dalszych badań nad molekularnymi mechanizmami regulującymi efektywność rozrodu u tego gatunku. Wartościowym elementem pracy jest również próba integracji wyników dotyczących określenia wpływu wieku dawczyń i warunków sezonowych na profil transkryptomiczny zarodków z wynikami analizy dekonwolucyjnej. Przedstawione w recenzji uwagi dotyczą zarówno kwestii redakcyjnych, jak i sposobu prezentacji wyników, spójności dokumentacji badawczej oraz opisu przeprowadzonych analiz bioinformatycznych i wymagają wyjaśnienia oraz doprecyzowania. Nie umniejszają one jednak wartości naukowej przeprowadzonych badań. Rozprawa opiera się na obszernym i wartościowym materiale badawczym, a zastosowane podejście eksperymentalne oraz zakres przeprowadzonych analiz świadczą o nakładzie pracy i zdolnościach badawczych Doktorantki.

Podsumowując, uważam, że przedstawiona do oceny rozprawa lek. wet. Klaudii Miętkiewskiej pt. „*Transcriptome analysis of early stages of bovine embryos obtained in vivo from Polish Holstein-Friesian cows*” spełnia wymagania stawiane rozprawom doktorskim określone w artyku 187 ustawy z dnia 20 lipca 2018 r. - Prawo o szkolnictwie wyższym i nauce (tj. Dz. U. z 2024 r. poz.1571 z późn. zm). Na tej podstawie zwracam się do Rady Dyscypliny Weterynarii Uniwersytetu Mikołaja Kopernika w Toruniu o dopuszczenie lek. wet. Klaudii Miętkiewskiej do dalszych etapów postępowania w sprawie nadania stopnia doktora nauk weterynaryjnych.

Z poważaniem

Agata Zimierska

