

ВІДГУК

офіційного опонента

на дисертаційну роботу

Лихенко Олександра Костянтиновича

"СПЕЦІАЛІЗОВАНА БАЗА ДАНИХ З ГЕННОЇ ЕКСПРЕСІЇ У МАТКОВО-ПЛАЦЕНТАРНОМУ КОМПЛЕКСІ ТА КРОВІ МАТЕРІ Й ДИТИНИ ТА ПЛАЦЕНТАРНИЙ ТРАНСКРИПТОМ УПРОДОВЖ ФІЗІОЛОГІЧНОГО ПЕРЕБІГУ ВАГІТНОСТІ",

представлену до захисту на здобуття наукового ступеня кандидата біологічних наук за спеціальністю 03.00.03 – молекулярна біологія

Порушення формування і функціонування плаценти є вагомим фактором ускладнень і патологій вагітності, як то, викидні, затримка росту плода, прееклампсія, тощо. Незважаючи на численні дослідження, етіологія та молекулярний аспект патогенезу цих ускладнень, більшою мірою залишаються незрозумілим. Це обмежує завчасну діагностику і перспективи таргетної терапії. Ключовий прорив у цьому напрямку є можливим лише з позиції комплексного аналізу накопичених дискретних даних, вирішують методи і підходи системної біології.

Головною метою дисертаційної роботи Лихенко Олександра Костянтиновича було дослідження особливостей змін плацентарного транскриптому впродовж нормального перебігу вагітності. Системно-біологічний підхід до вирішення цієї мети передбачав інтегративний аналіз експресії генів у плаценті людини. Відповідне дослідження було виконано шляхом сканування і комплексного аналізу оригінальної бази даних експресії генів, що була створена в межах актуального дисертаційного дослідження, і є його невід'ємною складовою. Для досягнення поставленої мети, Дисертантом було згенеровано похідну базу метаданих матково-плацентарного комплексу та крові матері й дитини. В основу зазначеної авторської бази даних, названої IGEA (Integrative Gene Expression Analysis) дисертантом було покладено результати критичної консолідації даних баз GEO і ArrayExpress, які містять як метадані, так і дані експресії генів.

Ревізія бази IGEA, дозволила відібрати зразки, відстежити однотипність метаданих, і інтегрувати відповідні показники експресії генів з баз GEO і ArrayExpress. Для вирішення поставлених завдань, вхідні дані експресії перетворювались на логарифмовані та нормалізовані матриці експресії, видалявся фактор серійності та інтегрувались дані експресії генів. Верифікація видалення ефекту серійності здійснювалася за методом головних компонент. Дисертантом було визначено кластери генів, що демонструють диференційну експресію протягом I/II та II/III триместрів вагітності, а також, відтворено мережі білок-білкових взаємодій їх продуктів. Виходячи з цих даних, Здобувачу вдалося визначити

біологічні процеси, на які збагачені гени окремих кластерів.

Все це дозволило Дисертанту визначити і охарактеризувати ключові гени збагачених біологічних процесів, а також, виявити ключові закономірності змін їхньої експресії у плаценті людини.

Зв'язок з науковими програмами, планами, темами. Робота виконана у межах тематики групи системної біології відділу ензимології білкового синтезу, що затверджена наказом від 26.11.2020р.№4, а також у рамках українсько-австрійського гранту 2023 - 2024 рр. № 0123U104257.

Наукова новизна дослідження та одержаних результатів, ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих у дисертації.

Дисертаційна робота Лихенко О.К. являє гарний приклад системно-біологічного дослідження, і була виконана на сучасному науковому та методичному рівні, ґрунтується на значному матеріалі спеціалізованих баз даних, оригінальних досліджень і авторських розробок. Усі експерименти були виконані Дисертантом особисто або за його безпосередньою участю і гармонічно інтегровані у тематику групи системної біології відділу ензимології білкового синтезу Інституту молекулярної біології і генетики. В межах представленого дослідження, Автором здійснено ревізію метаданих досліджень з експресії генів матково-плацентарного комплексу, крові матері й плода, а також, реалізовано аналіз транскриптому плаценти за умови нормального перебігу вагітності.

За результатами консолідації даних експресії генів, що були отримані з репозиторіїв GEO (Gene Expression Omnibus) та ArrayExpress, було сформовано в оригінальну похідну бібліотеку з максимально розширеним обсягом вибірки, що покращило статистичну вагомість наступного біоінформатичного аналізу. Із врахуванням системно-біологічних задач актуального дослідження, Здобувачем була виконана масштабна стандартизація даних експресії генів за антологіями MeSH і EFO. Таким чином, було отримано унікальний спеціалізований масив метаданих для матково-плацентарного комплексу і крові матері й плоду, який став ядром авторського репозиторію бази IGEA з веб-інтерфейсом, який дозволяє оптимізувати відбір метаданих за запитом користувача. Саме база IGEA стала ключовим об'єктом та інструментом віртуального скринінгу асамблей генів у здоровій тканині плаценти першого, другого та третього триместрів. Шляхом застосування розробленого Автором алгоритму інтегративного аналізу було отримано оригінальні дані, що характеризують якісні і кількісні зміни експресії генів між першим і другим, а також, другим і третім триместрами вагітності. Показано, що, перший часовий інтервал характеризується активними імунними і захисними реакціями, диференціацією клітин і ростом плаценти, прозапальними процесами, стабілізацією толерантності у контакті плодової частини плаценти і децидуальної тканини

матері. У свою чергу, показано що у другому інтервалі відбувається суттєвий спад прозапальних процесів, на заміну яким приходять протизапальні процеси, зменшення проліферативної активності і активності морфогенів, генів комітування, а також, активація метаболічних процесів, притаманних високодиференційованій тканині.

Вперше було відтворено профіль злomu експресії 24 генів: показано, що вектор їх експресії змінюється на протилежний у другому триместрі вагітності. Показано, що переломний момент в експресії зазначених генів синхронізовано із першим проявом клінічних симптомів. Висунуто припущення, про можливу залежність патологічних симптомів саме від зазначеної групи генів. Як аргумент на користь такого припущення Автор констатує наявність серед цих 24х генів ряду відомих маркерних генів, як то *LEP*, *IL6*, *CXCL8*, *HSDB2*, порушення експресії яких асоціюють з преєклампсією, найпоширенішим ускладненням при вагітності.

Автором вперше описано чотири універсальні траєкторії ко-експресії генів («Up-Up», «Up-Down», «Down-Down», «Down-Up») та запропоновано дві характерні схеми взаємодії диференційно експресованих генів у межах біологічного процесу, як то: взаємне виключення і синергія.

Загалом, аналіз матеріалів дисертації, автореферату, а також відповідних публікацій здобувача, дає мені підставу підтвердити наукову новизну і оригінальність одержаних результатів, а також, підтвердити обґрунтованість наукових положень, висновків і рекомендацій, сформульованих Дисертантом.

Практичне значення одержаних результатів. В результаті виконання дисертаційної роботи Автором було створено спеціалізовану базу IGEA (Integrative Gene Expression Analysis), яка узагальнює і стандартизує метадані отримані для зразків матково-плацентарного комплексу та крові матері й плоду, що надає можливість легкого і зручного відбору даних для вторинного аналізу експресії генів у зазначеної системі. За допомогою зазначеного інструменту, було узагальнено дискретні дані мікромасив-експериментів у плаценті людини опубліковані за останні 20 років, і сформовано інтегральний масив даних, що розкриває кількісні профілі експресії диференційно експресованих генів (ДЕГ), визначає функціонально збагачені біологічні процеси (ФЗБП) першого-другого (I/II) і другого-третього (II/III) триместрів вагітності. Вперше показано, що на межі першої і другої половини вагітності траєкторії експресії значної кількості генів «переламуються» (зокрема маркерний *LEP*), що збігається із часом появи клінічних симптомів преєклампсії. Автором описано чотири універсальні траєкторії ко-експресії генів («Up-Up», «Up-Down», «Down-Down», «Down-Up»), запропоновано три характерні схеми взаємодії диференційно експресованих генів у межах біологічного процесу, як то взаємне виключення та синергія. Узагальнені результати уточнюють молекулярний сценарій нормального розвитку плаценти, надається саме кількісна

характеристика змін у транскриптомі плаценти впродовж вагітності у нормі, що відкриває потенційну можливість майбутнього порівняння зі змінами транскриптому за умов патологій.

Особистий внесок здобувача

Викладені в дисертації результати було отримано Автором особисто або за його безпосередньої участі. Авторський внесок інших Авторів повністю розкрито у відповідних розділах дисертації і відображено у спільних публікаціях.

Повнота викладу результатів у наукових публікаціях.

Загалом, за матеріалами дисертації опубліковано 12 наукових праць: 5 статей у фахових наукових журналах та 7 публікації у матеріалах закордонних та вітчизняних наукових конференцій, симпозіумів і конгресів.

Структура дисертації

Дисертація складається з анотації, списку публікацій Дисертанта, вступу, огляду літератури, матеріалів і методів дослідження, результатів досліджень, обговорення та узагальнення отриманих результатів, 9 висновків, списку використаних джерел (121 найменувань) і 6 додатків. Робота викладена на 153 сторінках тексту, містить 32 рисунка і 3 таблиці.

Автором чітко сформульовано мету і завдання дослідження. Застосовані методи дослідження відповідають сучасному рівню і обґрунтовані завданнями дисертаційного дослідження. Висновки послідовні і відповідають меті і завданням дисертації.

Окремі дискусійні питання і зауваження до дисертації.

Опонент не має принципових зауважень до роботи. Декілька запитань для дискусії та зауважень технічного характеру наведено нижче:

1. Мені не сподобалось, що в «Меті дослідження» акцент зроблено на створенні бази даних. Дисертація захищається з біологічних наук і біологічна складова має бути на першому місці. Створення бази даних, веб-інтерфейсу, тощо, це у випадку спеціальності «Молекулярна біологія» лише програмні і алгоритмічні інструменти / етапи вирішення первинної молекулярно-біологічної задачі. Також, це зауваження відноситься і до самої назви дисертації.
2. Незважаючи на колосальний пласт наукової роботи і безсумнівну коректність підходів виконаного системно-біологічного дослідження, на жаль, слід зауважити що рукопис дисертації і автореферат перевантажені невдалими виразами, друкарськими помилками та програмістською термінологією, що, на жаль, псує загальне враження і ускладнює сприйняття результатів роботи.
3. Автор пише про те, що в наслідок об'єднання даних генної експресії з двох джерел (GEO, ArrayExpress) він отримав «єдиний якісний датасет» – масив первинних даних,

- мені це не зрозуміло. Мова скоріше за все йде про більш репрезентативну вибірку. Більш того, далі за текстом, Автор сам відзначає проблеми з неповними даними, дуплікацією інформації, тощо, які він був вимушений вирішувати для стандартизації інформації об'єднаної бібліотеки. Допускаю, що мова йде саме про репрезентативність.
4. Автором було застосовано метод головних компонент. Як відомо, це один з основних методів, що масштабування розмірності різноманіття вхідних даних до 2-3 компонент, що робить можливим їх кластеризації у 2D- і 3D- просторах. Мені незрозуміло, що в аспекті актуального дослідження Дисертант має на увазі у четвертому завданні дослідження: «Верифікувати видалення ефекту серійності методом головних компонент».
 5. «Наукова новизна» дисертації починається із «Створено й опубліковано на веб-сайті molecularbiology.ink базу IGEA,». Я розумію що база даних, це квінтесенція представленої роботи, але я думаю, що логічніше було розпочати з біологічного аспекту: «Виконано стандартизацію метаданих матково-плацентарного комплексу та крові матері й плоду... За результатом чого, було виконано узагальнення метаданих, що знайшло втілення у вигляді бази даних IGEA.»
 6. Базу IGEA створено за даними матково-плацентарного комплексу та крові матері й дитини, а подальший інтегративний аналіз виконується лише для зразків цілісної плаценти прешого дрогого і третього триместрів у нормі. Але на цьому недостатньо наголошується і створюється хибне враження, що інтегративний аналіз також виконувався не лише для зразків плаценти.
 7. Як я зрозумів з тексту дисертації (с. 38), з різних причин, декілька датасетів були виключені з аналізу, включаючи датасет E-GEOD-25906, якій не було включено у дослідження з причин технічної помилки завантаження даних. Не зовсім зрозуміло наскільки відсутність цієї інформації критична для фінального вмісту похідної бази даних, результатів кластеризації і чи планує Здобувач їх врахувати у майбутньому.

Висновок:

Незважаючи на численні зауваження, вважаю, що дисертаційна робота Лихенко Олександра Костянтиновича «Спеціалізована база даних з генної експресії у матково-плацентарному комплексі та крові матері й дитини та плацентарний транскриптом упродовж фізіологічного перебігу вагітності» представляє собою завершену наукову працю, присвячену важливій науковій і практичній проблемі. За структурою, актуальністю задач, теоретичним та методичним рівнем, новизною, фундаментальним та практичним значенням отриманих результатів дисертаційна робота цілком відповідає умовам п.11 «Порядку присудження наукових ступенів», затвердженого Постановою Кабінету Міністрів України від 24 липня

2013 р. №567, які пред'являються до кандидатських дисертацій, а її автор заслуговує на присудження наукового ступеня кандидата біологічних наук за спеціальністю 03.00.03 – молекулярна біологія.

Офіційний рецензент,

Завідувач лабораторії біоінформатики та структурної біології,
Відділу геноміки та молекулярної біотехнології,
ДУ «Інститут харчової біотехнології та геноміки
Національної академії наук України»
доктор біологічних наук,
старший науковий співробітник

Карпов П.А.

Підпис зав. лабораторії біоінформатики та структурної біології,
д.б.н., с.н.с. Карпова П.А.

ЗАСВІДЧУЮ

Вчений секретар

ДУ «Інститут харчової біотехнології
та геноміки НАН України»,
к.б.н., с.н.с.



Пірко Я.В.