

ІНСТИТУТ МОЛЕКУЛЯРНОЇ БІОЛОГІЇ ТА ГЕНЕТИКИ
НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ

ЛИХЕНКО ОЛЕКСАНДР КОСТЯНТИНОВИЧ



УДК 577.218

Спеціалізована база даних з генної експресії у матково-плацентарному комплексі та крові матері й дитини та плацентарний транскриптом упродовж фізіологічного перебігу вагітності

АВТОРЕФЕРАТ
дисертації на здобуття наукового ступеня
кандидата біологічних наук

03.00.03
молекулярна біологія

Київ – 2025

Дисертацією є рукопис.

Роботу виконано в групі системної біології відділу ензимології білкового синтезу Інституту молекулярної біології і генетики НАН України.

Науковий керівник:

професор, доктор біологічних наук
Оболенська Марія Юрївна,
Інститут молекулярної біології і генетики
НАН України, керівник групи системної
біології відділу ензимології білкового синтезу.

Офіційні опоненти:

доктор біологічних наук, професор
Сиволоб Андрій Володимирович,
ННЦ “Інститут біології та медицини”
Київського національного Університету імені
Тараса Шевченка

доктор біологічних наук, старший науковий
співробітник
Карпов Павло Андрійович,
завідувач лабораторії біоінформатики та
структурної біології
Інститут харчової біотехнології та геноміки
НАН України.

Захист дисертації відбудеться «30» **вересня 2025 р.** о 10.30 на засіданні спеціалізованої вченої ради Д 26.237.01 в Інституті молекулярної біології та генетики НАН України за адресою: 03680, м. Київ, вул. Академіка Заболотного, 150.

З дисертацією можна ознайомитися в бібліотеці Інституту молекулярної біології та генетики НАН України за адресою: 03680, м. Київ, вул. Академіка Заболотного, 150.

Автореферат розісланий «_» _____ 2025 р.

Вчений секретар
спеціалізованої вченої ради,

кандидат біологічних наук, с.н.с.



І.В. Крупська

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

- 1_2 — порівняння даних між першим і другим триместром вагітності
- 2_3 — порівняння даних між другим і третім триместром вагітності
- ArrayExpress — база даних функціональної геноміки
- CDF — файли анотації генів для платформи Affymetrix
- CEL — файли первинних даних мікрмасивів платформи Affymetrix
- FC — fold change, кратність зміни експресії гена
- FDR — false discovery rate (adjusted p-value) частка хибних відкриттів (скориговане p-значення)
- GEO — Gene Expression Omnibus, база даних
- GO — Gene Ontology, база даних
- GSE — GEO Series, ідентифікатор датасету в базі даних GEO
- IGEA — integrative gene expression analysis, база метаданих, розроблена автором даної роботи
- LogFC — logarithm fold change, логарифм кратності зміни (FC) за основою 2
- PCA — principal component analysis, аналіз головних компонент
- SQL — structured query language, структурована мова запитів до баз даних
- STRING — база даних білок-білкових взаємодій і її веб інтерфейс
- Датасет — структурований набір даних, який використовується для аналізу, машинного навчання або побудови гіпотез (англ. dataset)
- ДЕГ — диференційно експресований ген
- Ефект серійності — в аналізі даних ефект серії стосується систематичних, небіологічних відмінностей між серіями зразків, які можуть спотворити результати досліджень
- ПЗКМ — позаклітинний матрикс (англ. ECM, extracellular matrix)
- ФЗБП — функціонально збагачений біологічний процес (англ. functionally enriched biological process)

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Плацента — тимчасовий орган, що з першого дня вагітності забезпечує обмін продуктами життєдіяльності та імунний діалог між матір'ю та плодом. Порушення формування чи функції плаценти лежать в основі «великого акушерського синдрому» — сукупності ускладнень (викидні, затримка росту плода, прееклампсія тощо), які найбільше впливають на материнсько-дитячу захворюваність і смертність. Незважаючи на численні публікації, етіологія та молекулярний патогенез цих ускладнень залишаються не з'ясованими, що обмежує можливості ранньої діагностики й таргетної терапії (Kosińska-Kaczyńska, 2022; Yang C. et al. 2023).

Сучасні високопродуктивні методи дослідження сприяли накопиченню великих масивів транскриптомних даних плаценти у відкритих репозиторіях GEO та ArrayExpress. Однак більшість досліджень охоплює небагато зразків і ризняється за дизайном та анотацією, тож окремо не забезпечують належної статистичної сили. Інтегративний аналіз первинних мікроарей-даних на відміну від мета-аналізу дає змогу подолати цей бар'єр. Після стандартизації метаданих, нормалізації та корекції ефекту серійності об'єднання даних генної експресії з різних, але однотипних зразків плаценти, отримуємо єдиний якісний датасет, який дозволяє виявити гени, що статистично достовірно диференційно експресуються між триместрами та зрозуміти, як саме плацентарний транскриптом перебудовується протягом нормальної гестації. Таке знання формує міцний молекулярний ґрунт для подальших досліджень механізмів патологічних вагітностей і розробки біомаркерів для раннього прогнозування ускладнень і таргетної терапії.

Об'єктом дослідження є первинні дані зразків матково-плацентарного комплексу та крові матері й дитини, які знаходяться у відкритому доступі й охарактеризовані за генною експресією. **Предметом дослідження** є зміни в інтегрованому білок-кодуючому **транскриптомі плаценти між триместрами** впродовж фізіологічного перебігу вагітності.

Зв'язок роботи з науковими програмами, планами, темами. Роботу виконано відповідно до 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.

Мета роботи — створити спеціалізовану базу метаданих для зразків матково-плацентарного комплексу та крові матері й дитини, які знаходяться у відкритому доступі й охарактеризовані за генною експресією, і визначити зміни у плацентарному транскриптомі впродовж нормального перебігу вагітності у відібраних і об'єднаних зразках методом інтегративного аналізу.

Для досягнення поставленої мети було сформульовано завдання дослідження:

1. Створити вторинну спеціалізовану базу метаданих для зразків матково-плацентарного комплексу та крові матері й дитини,

- використовуючи бази GEO і ArrayExpress, які містять як метадані, так і дані генної експресії.
2. Відібрати з бази IGEA однотипні за метаданими зразки і отримати з баз GEO і ArrayExpress відповідні дані генної експресії за ідентифікаційними номерами досліджень.
 3. Перетворити первинні дані експресії (дані про інтенсивність випромінювання на пробах мікромасивів) у логарифмовані та нормалізовані матриці експресії, видалити ефект серійності та провести інтеграцію даних генної експресії.
 4. Верифікувати видалення ефекту серійності методом головних компонент.
 5. Визначити диференційно експресовані гени (ДЕГи) між I і II та II і III триместрами вагітності.
 6. За даними з бази STRING побудувати відповідні ДЕГам мережі міжбілкових взаємодій і асоціацій та кластеризувати вузли (білки) мереж.
 7. Визначити біологічні процеси, на які збагачені гени у кожному з кластерів.
 8. Охарактеризувати функціонально збагачені на біологічні процеси гени і виявити закономірності змін в експресії генів у межах цих процесів.

У дослідженні застосовано комплекс **методів** біоінформатичного аналізу: *збір і стандартизація даних* (бази даних GEO, ArrayExpress; категоризація метаданих за онтологіями MeSH і EFO); програмні засоби Python/Django/Postgres для розробки веб-бази IGEA; *обробка даних експресії* (RMA в R пакеті affy для .CEL-файлів Affymetrix; фон-корекція, log₂-трансформація й квантильна нормалізація в limma для Illumina); *інтеграція даних* (емпіричний метод Байеса реалізований у методі ComBat з R пакета sva); *контроль якості* первинної обробки даних (PCA, t-SNE); *визначення диференційно експресованих генів* (t-тест, лінійні моделі logFC з FDR-корекцією); *кластерний аналіз* (ієрархічна кластеризація за методом «fast-greedy» на графі білок-білкових взаємодій STRING); *визначення функціонально збагачених груп генів* (гіпергеометричний тест, за даними з баз STRING і Gene Ontology); побудова теплокарт близькості біологічних процесів (індекс Джаккарда, R пакет pheatmap).

Наукова новизна. Створено й опубліковано на веб-сайті molecularbiology.link базу IGEA, яка надає стандартизовані метадані для зразків матково-плацентарного комплексу та крові матері й плоду і надає можливість для проведення вторинного аналізу генної експресії. За допомогою створеного інструменту виконано інтегративне об'єднання даних мікромасивів за 20 останніх років, отримано новий узагальнений датасет і вперше надана кількісна характеристика рівня експресії достовірних ДЕГів та встановлено біологічні процеси, на які функціонально збагачені ці ДЕГи, для інтервалів перший-другий (1_2) та другий-третій триместри (2_3); показано, що на межі першої і

другої половини вагітності траєкторії експресії численних генів «переламуються» (зокрема маркерний LEP), що збігається з часом появи клінічних симптомів преєклампсії. Описано чотири універсальні траєкторії ко-експресії генів («Up-Up», «Up-Down», «Down-Down», «Down-Up») та три характерні схеми взаємодії ДЕГів у межах біологічного процесу: взаємне виключення (наприклад, прозапальні хемокіни проти гомеостатичних), синергія (активація активаторів морфогенезу з одночасним інгібуванням їх антагоністів та активація інгібінів призводить до специфічного інгібування активінів і пригнічення морфогенетичних процесів). Разом ці спостереження уточнюють молекулярний сценарій нормального розвитку плаценти.

Особистий внесок автора в роботу був визначальним на етапах відбору датасетів, стандартизації метаданих, створення бази даних і веб-сайту з доступом до неї, біоінформатичного аналізу даних генної експресії, і суттєвим — на етапі біологічної інтерпретації диференційної експресії генів. Автор вдячний науковому керівникові Оболенській М.Ю за супровід на всіх етапах виконання роботи та, зокрема: за допомогу в підборі вдалих термінів з онтологій EFO і MESH при створенні бази метаданих IGEA (публікація «Creation of gene expression database on preeclampsia-affected human placenta»); за додаткову верифікацію коректності роботи пайплайну інтегративного аналізу при його розробці шляхом оцінки, чи мають ті чи інші результати, які видає пайплайн, біологічний зміст (публікація «Consecutive integration of available microarray data for analysis of differential gene expression in human placenta»); за значну допомогу в біологічній інтерпретації встановленої диференційної експресії генів та у виявленні цікавих закономірностей змін генної експресії в межах різних біологічних процесів (публікації «Changes in the human placental transcriptome during the physiological course of pregnancy» та «The knowledge of chromosomal sex is important for large-scale analysis of gene expression») та за численні правки та рекомендації при підготовці публікацій до друку. Автор висловлює подяку м.н.с. Фроловій А. О. за перевірку коректності застосованих в усіх згаданих публікаціях біоінформатичних методів. Автор також вдячний С. А. Фаундс та іншим авторам датасетів, які в ході особистого листування поповнили наявні дані додатковими неопублікованими метаданими.

Апробація матеріалів дисертації проводилася на наукових заходах: 17–20 жовтня 2017 р. на 2017 IEEE International Young Scientists Forum on Applied Physics and Engineering (YSF-2017), м. Львів, Україна EPS Young Minds; 30 жовтня – 3 листопада 2017 р. на China–Ukraine International Symposium on Innovation and Technology, Shandong Academy of Sciences, Циндао, Китай; 26 травня 2021 р. на XV Всеукраїнській конференції молодих вчених з міжнародною участю, Інститут молекулярної біології і генетики НАН України, Київ; 19 травня 2023 р. на «Біотехнологія XXI століття»: XVII Міжнародній науково-практичній конференції, КПІ ім. Ігоря Сікорського, Київ; 19–21 листопада 2024 р. на Міжнародній конференції з нейронаук та наукових читаннях, Київ; 17–20 вересня 2024 р. на International Scientific Conference

BioGENext, Київ, Україна. Ці виступи були публічною апробацією підходів та результатів дослідження серед фахової спільноти.

Публікації. За темою дисертації опубліковано 5 наукових публікацій: дві статті індексовано у міжнародній наукометричній базі даних Scopus (Q4), одну статтю представлено у виданнях, що входять до інших міжнародних наукометричних баз даних, дві статті опубліковано у наукових фахових виданнях України; 7 тез доповідей на міжнародних та всеукраїнських науково-практичних конференціях, симпозіумах і конгресах.

Структура та обсяг роботи. Дисертація складається з анотації, переліку публікацій здобувача, вступу, огляду літератури, матеріалів і методів дослідження, результатів інтегративного біоінформатичного аналізу, розділу з обговоренням та узагальненням отриманих результатів, висновків, списку використаних джерел і додатків. Роботу викладено на 153 сторінках тексту; містить 32 рисунка, 3 таблиці та 6 додатків.

ОСНОВНИЙ ЗМІСТ

Літературний огляд дисертації містить сучасні уявлення про плаценту і її роль у триєдиній системі «мати–плацента–плід». Проаналізовано будову (цитотрофобласт, синцитіотрофобласт, ворсини хоріона, гемоплацентарний бар'єр), функції (транспортування газів і метаболітів, синтез гормонів — hCG, PL, регуляторних факторів) та етапи розвитку (інвазія трофобласта, ремодельовання маткових судин, диференціювання, морфогенез і ріст) плаценти. Окремі розділи присвячено впливу статі плоду на формування і функціонування плаценти і механізмам дисфункції плаценти, що призводять до ускладнень вагітності (пreekлампсія, затримка внутрішньоутробного росту).

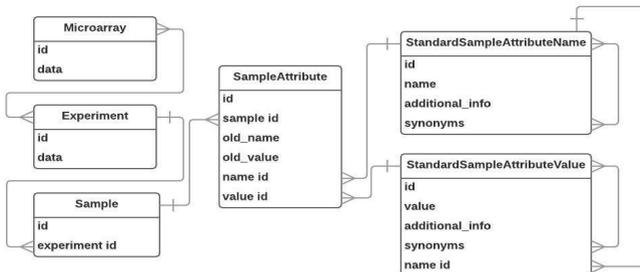
Другий блок огляду містить характеристику періодів вагітності та концепцію «вікон вразливості» і огляд системних підходів до вивчення молекулярних процесів у плаценті за допомогою високопродуктивних multi-omics технологій. Описано методи інтегративного аналізу даних генної експресії (мета- і інтегративний аналіз початкових даних), підходи до нормалізації й корекції batch-ефекту (зокрема, функція ComBat), а також обґрунтовано необхідність створення спеціалізованої бази стандартизованих метаданих для уможливлення інтегративного аналізу даних генної експресії.

Матеріали та методи досліджень. Матеріалом для виконання дисертаційної роботи слугували первинні дані для зразків матково-плацентарного комплексу і крові матері й дитини з охарактеризованою генною експресією, які було отримано з баз ArrayExpress і GEO. У дослідженні застосовано комплекс методів біоінформатичного аналізу. Через застосування *інформаційного пошуку* відібрали з відкритих баз зразки матково-плацентарного комплексу, для яких дані генної експресії отримані мікроарей-технологією. Метадані стандартизували вручну за MeSH- та EFO-термінами; власну базу IGEA розроблено з використанням середовища розробки Python +

Django + Postgres; база IGEA доступна за посиланням <https://molecularbiology.ink/igea>. Початкові дані обробляли окремо для платформ *Affymetrix* (RMA-процедура пакета *affy* з актуальними Brainarray CDF, агрегація проб, фон-корекція, квантильна нормалізація, \log_2) і *Illumina* (фон-корекція, \log_2 та квантильна нормалізація у *limma*). Для кожного гена залишали зонд із найвищою середньою інтенсивністю значення експресії. Об'єднану матрицю формували усуненням batch-ефекту методом ComBat (*sva*). Якість нормалізації перевіряли PCA та t-SNE. Диференційно експресовані гени (ДЕГи) в інтервалах 1_2 та 2_3 знаходили лінійними моделями (t-тест, \log_{FC} , FDR). Біологічну інтерпретацію ДЕГів отримали у кілька кроків: списки ДЕГів проектували на мережу білок-білкових асоціацій STRING і об'єднували вузли алгоритмом fast-greedy, щоб виділити компактні функціональні кластери. Далі для кожного з таких кластерів за допомогою Fisher-тесту визначали надрепрезентацію термінів Gene Ontology та шляхів, анотованих у STRING, тобто визначали, на які біологічні процеси гени у кластерах є збагаченими (в тексті – ФЗБП, функціонально збагачені біологічні процеси). Після цього за допомогою R пакета *heatmap* будували теплові карти близькості біологічних процесів за кількістю спільних генів, що до них входять, беручи індекс Джаккарда за критерій близькості.

Результати досліджень

Структура бази даних IGEA. На рис. 1 позначено схему реляційної бази IGEA (Integrative Gene Expression Analysis). Прямокутники – таблиці бази, стрілки – зв'язки між таблицями. Структура бази дозволяє: зберігати початкові назви й значення характеристик, а також їхні стандартизовані відповідники згідно з онтологіями MeSH та EFO; зберігати зразки з різними наборами



характеристик і додавати нові характеристики без перебудови структури бази.

Рис. 1.
Структура реляційної бази даних IGEA

У базу IGEA включено 1202 зразки матково-плацентарного комплексу і зразки крові матері й дитини. Зразки класифіковані за трьома основними критеріями — діагнозом, строком гестації та типом біологічного матеріалу (рис. 2). Тип біологічного матеріалу включає категорії: цілісна плацента,

базальна пластина, децидуальна тканина, ворсини хоріона, окремі клітинні популяції плаценти, кров матері та пуповинна кров.



Рис. 2. Схема класифікації зразків матково-плацентарного комплексу за трьома ключовими атрибутами — діагнозом, строком гестації та типом біоматеріалу

Структура даних генної експресії. З бази даних IGEA було відібрано 8 датасетів генної експресії в тканині плаценти з трьох триместрів за умов нормального перебігу вагітності (всього 77 зразків: 13 першого, 9 другого й 55 третього триместру) і проведено первинну обробку даних з верифікацією якості обробки (рис. 3) і попарну інтеграцію даних з першого і другого і другого і третього триместрів вагітності як зазначено в розділі Матеріали і методи.

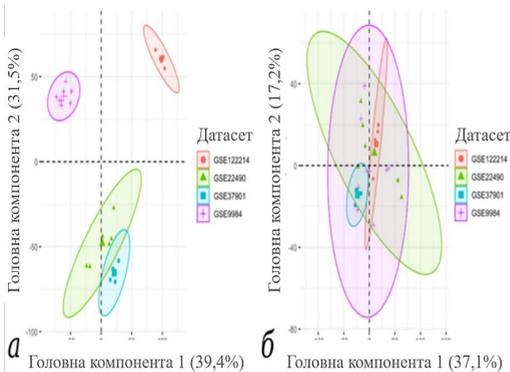


Рис. 3. Результати аналізу головних компонент до (3а) й після (3б) видалення ефекту серійності в зразках з першого і другого триместрів. Найбільшу варіативність (головні компоненти dim1 та dim2) вносять різні датасети (а), і варіативність зникає після видалення ефекту серійності (б)

Загальна характеристика диференційно експресованих генів між триместрами. Диференційно експресовані гени визначено між першим і другим (253 ДЕГи) і другим і третім (489 ДЕГВ) триместрами і проведено їх

кластеризацію. Кластери названі відповідно до найбільш представлених у них біологічних процесів (рис. 4).

Кластери в інтервалі 1_2 мають покриття збагачення від 0,87 до 1,0, а кластери в інтервалі 2_3 від 0,91 до 0,93 крім кластерів 3 і 5 з відповідними значеннями 0,14 і 0,57. Покриття збагачення — величина в межах 0 – 1, яка відображає частку генів, які належать (покриваються) до принаймні одного функціонально збагаченого процесу в процесу за даними програми STRINGdb. Тобто переважна більшість генів у більшості кластерів належить до функціонально збагачених процесів за даними програми STRINGdb.

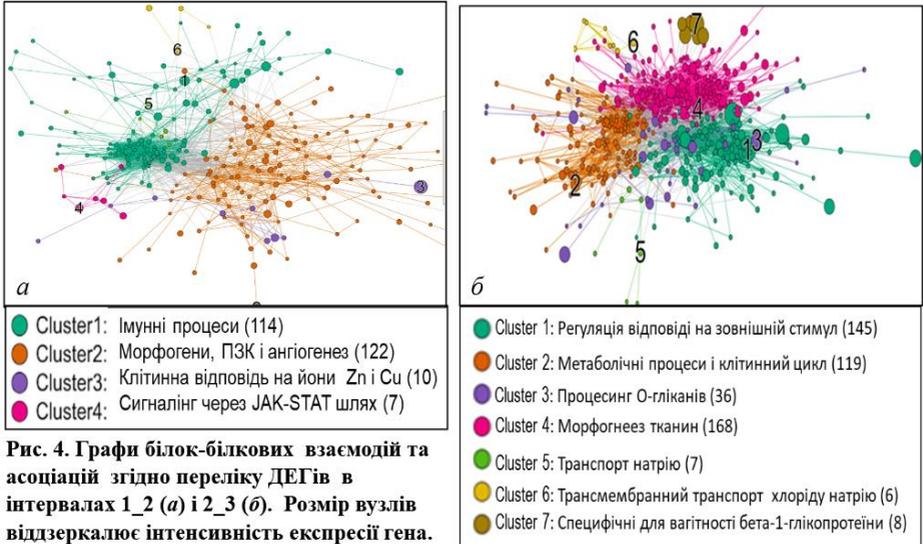


Рис. 4. Графи блоків-білкових взаємодій та асоціацій згідно переліку ДЕГів в інтервалах 1_2 (а) і 2_3 (б). Розмір вузлів віддзеркалює інтенсивність експресії гена.

Кластери диференційно експресованих генів і функціонально збагачені біологічні процеси між першим і другим триместром (порівняння 1_2). Тут і надалі ДЕГи позначаються жирним курсивом, а logFC (логарифм кратності зміни експресії за основою 2) жирним прямим шрифтом в дужках, наприклад, *CCL2* (1,64). Відповідні білки позначені жирним прямим шрифтом, наприклад, **CCL2**. В кожному кластері визначали функціонально збагачені біологічні процеси (ФЗБП) і ДЕГи, які відносяться до цих процесів. На рисунках в кожному кластері представлені ФЗБП з найбільшим числом генів (топ) і ДЕГи, які належать до одного з ФЗБП. Ілюстрації ФЗБП і ДЕГів надані у вигляді блокових діаграм. На діаграмах ДЕГів довжина блоків і їх спрямованість від нульового значення означають у скільки разів і в який бік, збільшення або зменшення, змінився рівень експресії відповідного гена у міру розвитку плаценти, тобто другий триместр проти першого і третій проти другого.

Кластер 1. Імунні процеси. Вагітність є викликом для імунної системи матері, оскільки плод, який розвивається всередині генетично відмінного організму матері, не має бути відторгнутим і одночасно має бути захищеним від дії патогенних чинників, які можуть потрапити з крові та матки матері (Zhu L. et al., 2018; Ding et al. 2022). Імунні процеси включають активацію системи комплементу і прозапальних цитокинів (рис. 5), перестройку ПЗКМу і активне переміщення імунних клітин до вогнищ тимчасового запалення.

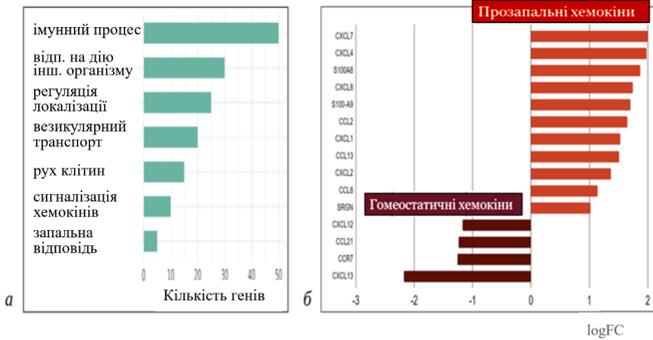


Рис. 5. Топ-7 ФЗБП (а) і гени прозапальних (червоний колір) і гомеостатичних (коричневий колір) хемокинів (б) у другому триместрі у порівнянні з першим

Кластер 2. Морфогени, позаклітинний матрикс і ангиогенез. Для даного кластера STRINGdb визначає ФЗБП, як такі, що пов'язані з розвитком анатомічних структур ембріона, легень, піднебіння, кінцівок і т.п. (рис. 6а). Скоріше за все це пояснюється наявністю в плаценті стовбурових клітин, які у різному оточенні можуть розвиватися в бік формування структур плода або структур плаценти.

Генну експресію в кластері проілюстровано генами ключових висококонсервативних морфогенних каскадів BMP-SMAD, Wnt, Notch та Hippo, гомеодоменими транскрипційними факторами, які задають просторову полярність і напрямок диференціювання (комітування) клітин. Звертає увагу той факт, що експресія генів, які активують морфогенетичні процеси (цілком забарвлені блоки), підвищується, а тих, які інгібують процес (забарвлений контур), знижується, сприяючи оптимізації процесу (рис. 6б).

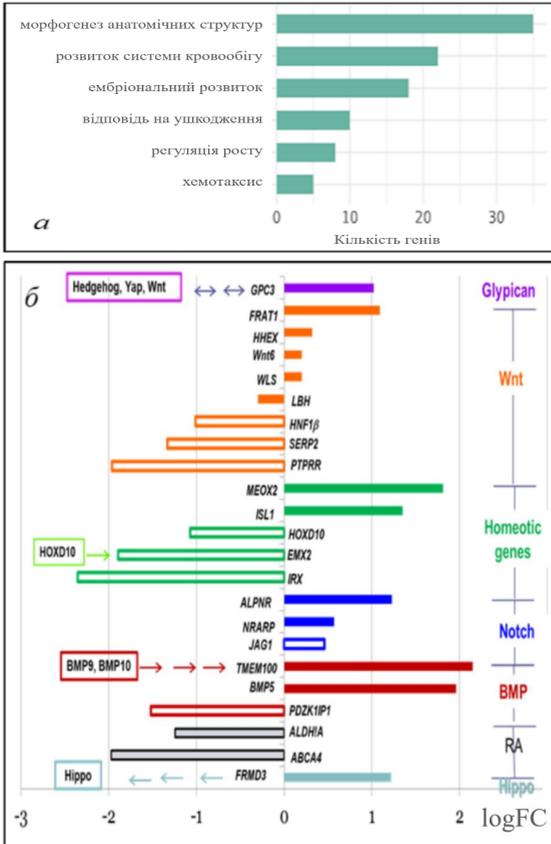


Рис. 6. Топ-6 ФЗБП (а) і диференційно експресовані гени у кластері «Морфо-гени, ПЗКМ і ангиогенез» (б). Примітка: Glurican-3 (GPC3) – протеоглікан клітинної поверхні, взаємодіє з численними лігандами і морфогенами; RA - ретинова кислота, важлива на найранішому етапі розвитку і небезпечна в подальшому. Стрілками показані напрямки, за яким ген-регулятор впливає на мішень.

Кластер 3. Клітинна відповідь на іони цинку й міді. Кластер вирізняється зростаючою активністю генів, які беруть участь у метаболізмі жирів (**PDK4**, домінуючий регулятор метаболізму глюкози та жирних кислот) (Yu et al., 2018). Одночасно знижується активність металотіонеїнів (**MTI**), які хелатують іони важких металів і зв'язують вільні радикали, і глутатіонпероксидази (**GPX3**), яка відновлює органічні пероксиди і перекис водню. Обидва зниження позбавляють захисту від оксидативного стресу на цьому етапі (Yang R. et al., 2024). Ген **PAEP (-4,43)** інгібує активність імунних клітин на етапі імплантації і забезпечує прикріплення бластоцисти до стінки матки.

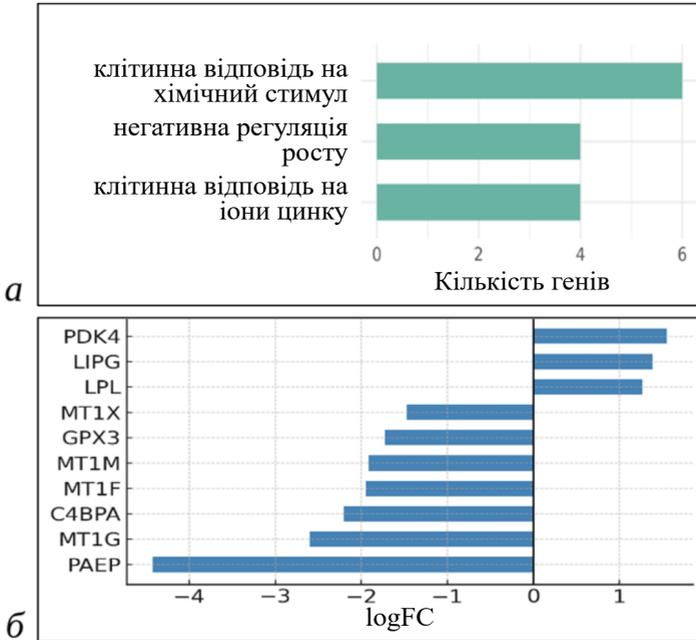


Рис. 7. Топ-3 функціонально збагачених біологічних процесів (а) і диференційно експресовані ДЕГи (а) у кластері «Клітинна відповідь на іони цинку й міді»

Кластер 4. Сигналінг через JAK-STAT шлях. Члени цього кластера беруть участь у сигнальних шляхах інтерлейкіну-35 та JAK-STAT, опосередковуючи активність цитокінів родини ІЛ-6/ІЛ-2 і факторів росту (рис. 8). Дія ростових гормонів перегукується з дією іншого білкового гормону, лептину (*LEP*, -2,38). Основна функція лептину — регуляція запасів жиру й сигналізація про відчуття голоду/ситості. Припускається, що зниження експресії гена лептину від I до II триместру може пояснювати, зокрема, підвищення апетиту на початку вагітності.

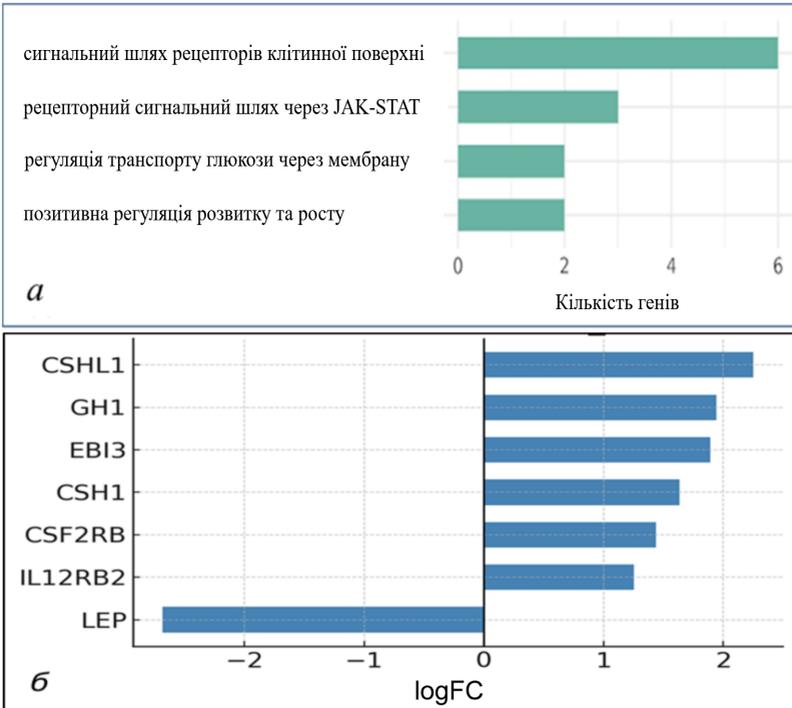


Рис. 8. Топ-4 функціонально збагачені біологічні процеси (а) і диференційно експресовані гени (б) у кластері «Сигналінг через JAK-STAT шлях»

Кластери диференційно експресованих генів й функціонально збагачені біологічні процеси між другим і третім триместрами (порівняння 2_3). Період 2_3 у порівнянні з 1_2 інтервалом характеризується більшою кількістю кластерів диференційно експресованих генів. Функціонально збагачені біологічні процеси являють більшу різноманітність і часто іншу спрямованість одних і тих же процесів у порівнянні з першою половиною вагітності.

Кластер 1. Регуляція відповіді на зовнішній стимул. Цей кластер об'єднує гени, які засвідчують карколомні зміни, яких зазнають ФЗБП і експресія диференційно експресованих генів (рис. 9а). На відміну від першої половини вагітності знижується експресія хемокинів та інтерлейкінів (рис. 9б). До цих змін додається нерівномірне зниження експресії генів системи комплементу (*C3*, -1,24; *C7*, 1,07; *CFD*, -1,05).

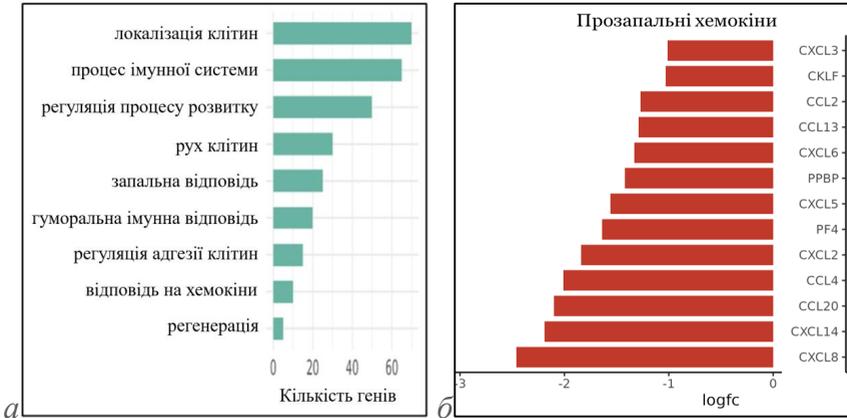


Рис. 9. Топ-9 функціонально збагачених біологічних процесів (а) та зміна експресії хемокінів між II і III триместром вагітності (б) у кластері «Регуляція відповіді на зовнішній стимул»

У цей період у плаценті відбувається перебудова ліпідного обміну: спостерігається одночасне зниження експресії аполіпопротеїнів, які плацента синтезувала і секретувала в першій половині вагітності для забезпечення себе і плоду ліпідами з крові матері. У другій половині синтетичні процеси перебирає на себе плод. Внутрішньоклітинний **FABP4 (5,45)** зв'язує жирні кислоти та спрямовує їх до ядерного рецептора і транскрипційного фактора PPAR γ , який регулює метаболізм ліпідів і глюкози, диференціацію клітин та імунітет. Активація генів **SCARB1 (1,8)** і **CETP (2,16)** свідчить про посилений холестериновий обмін у плаценті, необхідний для стероїдогенезу та розвитку мозку плоду. Ескалація експресії **CRH (6,14)** запускає гіпоталамо-гіпофізарно-надниркову вісь, готуючи організм матері до пологів і підтримуючи нейророзвиток плоду. Нарешті, змінюється тип плацентарного гемоглобіну: знижується фетальний варіант і підвищується рівень субодиноці, характерної для мінорної форми гемоглобіну дорослої людини.

Кластер 2. Метаболічні процеси і клітинний цикл. У кластері Метаболічні процеси і клітинний цикл найбільш представлені процеси, які відповідають назві кластеру (рис. 10).

Експресія генів, прямо чи опосередковано причетних до фаз клітинного циклу, достовірно зменшується між II і III триместром в 2–4 рази за винятком гена **KIF2A (1,07)** (рис. 10б). Білок **KIF2A** необхідний для нормального шиккування хромосом у метафазній пластинці і функціонування веретена поділу під час мітозу. Пригнічення експресії генів клітинного циклу підтверджує загальне суттєве зниження проліферації клітин у плаценті в міру прогресування гестації.

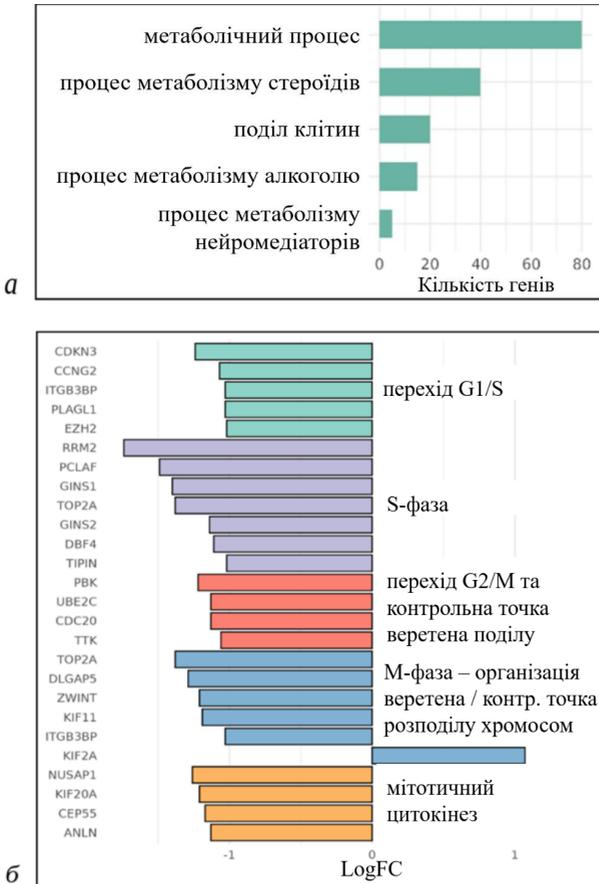


Рис. 10. Топ-5 функціонально збагачених біологічних процесів (а) та зміна експресії генів клітинного циклу (б) у кластері «Метаболічні процеси і клітинний цикл»

На цьому етапі вирізняється активація генів біосинтезу стероїдів (*FDX1*, 1,23; *CYP11A1*, 2,02; *INHBA*, 3,58; *HSD17B1*, 1,22; *HSD3B1*, 2,28; *CYP19A1*, 2,42; *HSD3B2*, -1,71), що корелює з даними ЗТ-ПЛР (Karahoda et al., 2021) та інгібування генів синтезу холестерину (*CYP51A1*, -1,14; *MSMO1*, -1,67; *HMGCS1*, -1,03 і *DHCR24*, -1,24). Якщо в першій половині плацента сама забезпечувала потреби

свої і плоду у холестерині, то у другій вона активно транспортує материнський холестерин, який є в достатку у материнській крові на даному етапі (Li. Y. et al., 2024). Синтез холестерину і аполіпропротеїнів перебирає на себе плод, а використання холестерину для стероїдогенезу відбувається на високому рівні.

Кластер 3. Процесинг О-гліканів. Кластер містить гени, які відповідають за: формування і секрецію муцинів, О-глікозилювання, клітинну адгезію та деякі функції, пов'язані з імунітетом. О-глікозилювання — ключовий і єдиний ФЗБП у цьому кластері. Два найвиразніших ДЕГи — *HLA-G* (-2,13) і *ALPP* (6,65) — є маркерами позаворсинкового трофобласту та зрілого синцитіотрофобласту, відповідно. *HLA-G* (лейкоцитарний антиген-С людини) задіяний в імунній толерантності, а *ALPP* у сприянні транспортуванню поживних речовин через матково-плацентарний бар'єр.

(Goldman-Wohl et al., 2000; Reiswich et al., 2021). Алгоритм кластеризації fast-greedy об'єднав у цей кластер біологічно розрізнені гени - «одинаки», які мають низький рівень взаємодії як між собою (згідно з базою даних STRING), так і з іншими елементами мережі.

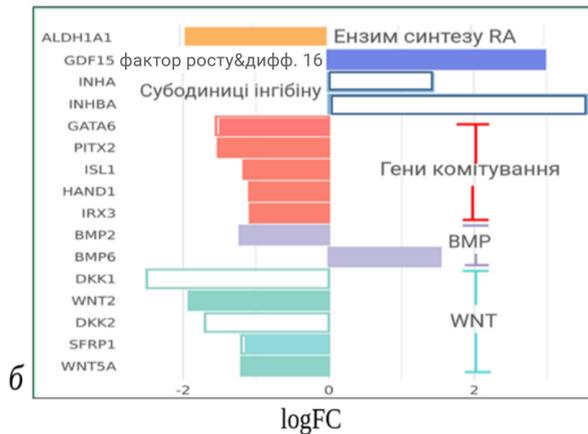
Кластер 4. Морфогенез тканин і фактори комітування.



У цьому кластері спостерігається сповільнення Wnt-керованого морфогенезу (рис. 11).

Рис. 11. Топ-5 ФЗБП (а) та експресія генів, пов'язаних з морфогенезом і комітуванням, у третьому триместрі у порівнянні з другим (б).

Примітка: гени-активатори позначені зафарбованими блоками; антагоністи Wnt сигналіngu, *DKK1* і *DKK2*, і гени *INHA* і *INHBA* білків-інгібіторів позначені незафарбованими блоками.



Як канонічні та неканонічні гени-ліганди WNT, так і їхні антагоністи (*DKK*, *SFRP*) знижують експресію, що

свідчить про пригнічення шляху в цілому (Xue C. et al., 2025). Те саме стосується представників надродина TGF- β , до якої належать *BMP2*, інгібіни А та В (*IHA* і *INHBA*), які послаблюють активін-залежну диференціацію клітин. *BMP6* і *GDF15* складають виключення із загального переліку генів, експресія яких знижується. *BMP6* задіяний в імплантації (Deng J, Li Y., 2021) і ймовірно у стероїдогенезі і виживанні клітин наприкінці вагітності; підйом секреції плацентарного *GDF15* асоціюють з ранніми симптомами нудоти й гіперемезису, а його підйому наприкінці вагітності поки немає функціонального пояснення. Іншу групу складають гени комітування (*GATA6*, *HAND1*, *ISL1* та *PITX2*). Вони кодують транскрипційні фактори, які активні в ембріогенезі і відповідають за формування органів і структур організму, осей тіла, ліво-праву симетрію (Lynch et al., 2025). В плаценті на ранньому етапі

вони регулюють закладку алантоїсу/пуповини, хоріон-алантоїсної пластинки й судинної системи плаценти; проліферацію і диференціювання клітин. У другій половині вагітності ближче до закінчення вагітності роль перелічених факторів вичерпана, і експресія генів знижується.

До цього кластера потрапили також гени, які відповідають за синтез колагену і гіалуринової кислоти. Ці компоненти позаклітинного матриксу асоціюють з процесами розвитку (Reddi, 1976). Не виключено, що зменшення щільності структури через зниження колагену є необхідним передумовою до майбутніх пологів (Shi et al., 2020).

Кластер 5 «Транспорт натрію» і кластер 6 «Трансмембранний транспорт хлориду» (рис 12). Кластер 5 об'єднує гени salvage-шляху синтезу нуклеозидів і переносників Na^+ . Такий шлях притаманний диференційованим клітинам з малою проліферативною активністю і відповідає стану плаценти наприкінці гестації (Rodriguez et al., 2021).

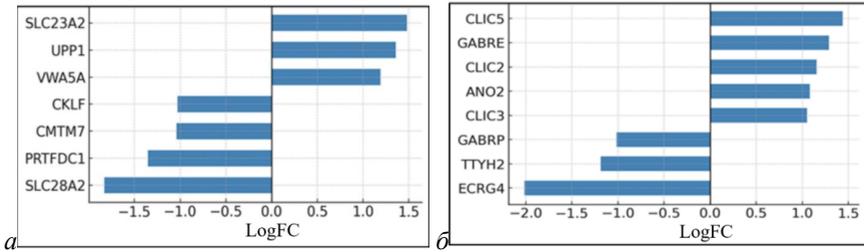


Рис. 12. Диференційна експресія генів у кластері 5 «Транспорт натрію» (а) і кластері 6 «Трансмембранний транспорт хлориду» (б) у третьому триместрі у порівнянні з другим

Гени транспорту Cl^- та ширша категорія «транспорт» зосереджені у кластері 6: хлорид створює осмотичний градієнт для руху води через синцитіотрофобласт (Doughty et al., 1998) і, проходячи крізь канали, регулює мембранний потенціал, що впливає на секрецію гормонів та обмін інших іонів (Riquelme, 2009). Диференційно експресовані гени обох кластерів показано.

Кластер 7. Специфічні для вагітності бета-1-глікопротеїни (pregnancy-specific glycoproteins, PSG) — це білки з ковалентно зв'язаними олігосахаридами, приєднаними до бокових ланцюгів амінокислот. Глікопротеїни є мажорними білками в кровотоці матері на пізніх термінах вагітності. Десять генів *PSG* кодують білки (PSG1 — PSG9, PSG11). Відомо, що ці білки мають імунорегуляторну, проангіогенну та антитромбоцитарну функції. Вони взаємодіють і активують протизапальні цитокіни, фактори росту TGF β 1 та TGF β 2, що робить їх одними з небагатьох відомих біологічних активаторів цих важливих цитокінів. Між другим і третім триместром суттєво (у 8 — 32 рази) підвищується експресія генів *PSG* 1, 2, 3, 4, 9 і 11 (рис. 13).

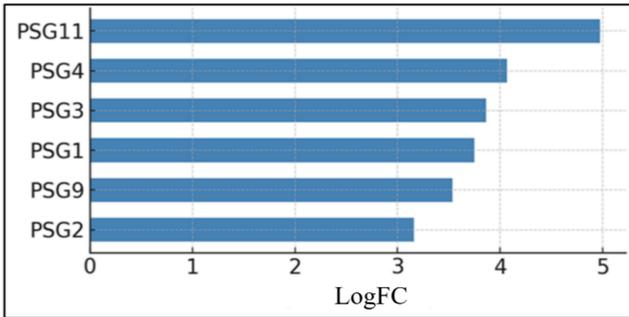


Рис. 13. Експресія генів, які кодують специфічні для вагітності бета-1-глікопротеїнів (кластер 7), в третьому триместрі у порівнянні з другим

Порівняльний аналіз генної експресії у плаценті першої і другої половини вагітності (порівняння між 1_2 та 2_3) виявив спільні 54 диференційно експресовані гени (рис. 14), які за характером змін в процесі вагітності розділяються на чотири групи: UpUp і DownDown, в яких експресія неухильно зростає або падає, відповідно; UpDown і DownUp, в яких спрямованість експресії змінюється у другому триместрі на відміну від першого триместру.

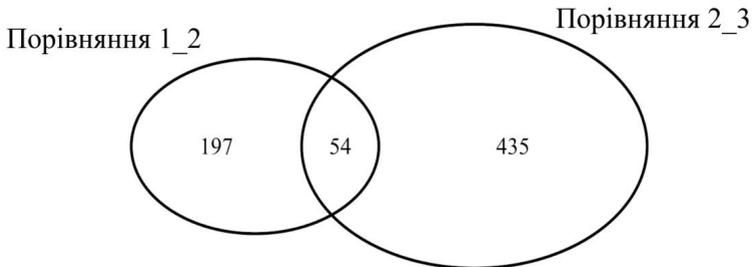


Рис. 14. Діаграма Венна: спільні диференційно експресовані гени з $|\logFC| > 1$ у порівняннях 1_2 і 2_3

UpUp (рис. 15). Експресія вісімнадцяти генів наростає як у 1_2, так і у 2_3. Вони позначають перехід плаценти до зрілих функцій — від антимікробного захисту й згортання крові до ендокринної перебудови та переключення метаболізму на жирні кислоти.

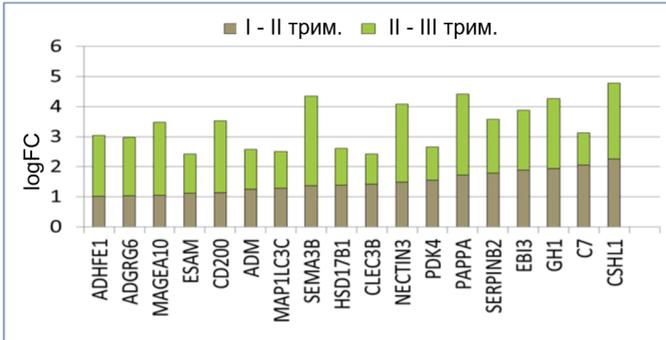


Рис. 15. Експресія спільних диференційно експресованих генів з групи UpUp у першій і другій половині вагітності

UpDown (рис. 16). Експресія двадцяти одного гена підвищується у 1_2, але знижується у 2_3. Блок хемокінів, простагландин-синтетаза і цитокіни сигналізують про ранній прозапальний/ангіогенний «спалах», що потрібен для імплантації та початкового ремоделювання артерій; у 2_3 ця реакція стихає, водночас активується ангіотензинова система та вивільняється PAPP.

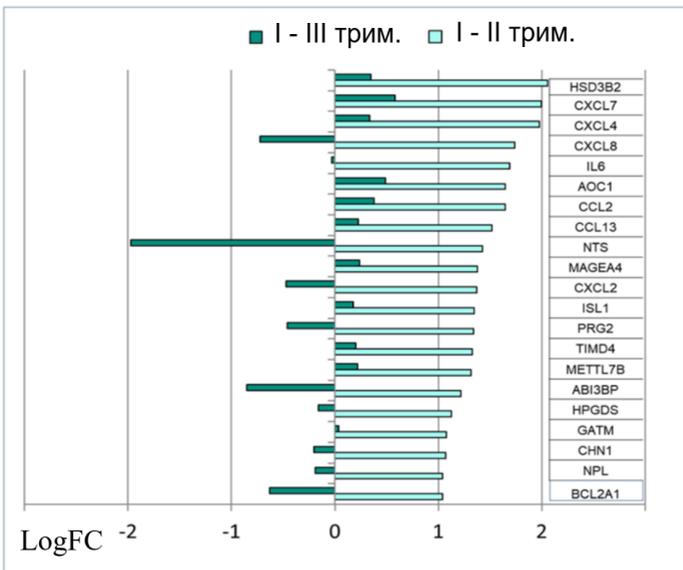


Рис. 16. Експресія диференційно експресованих генів з групи UpDown у першій і другій половині вагітності

DownDown (рис. 17). Експресія дванадцяти генів згасає у 1_2 і продовжує згасати у 2_3. Це регулятори імплантації, фетального гемоглобіну й раннього матричного ремоделювання, чия роль завершується ще до періоду 2_3.

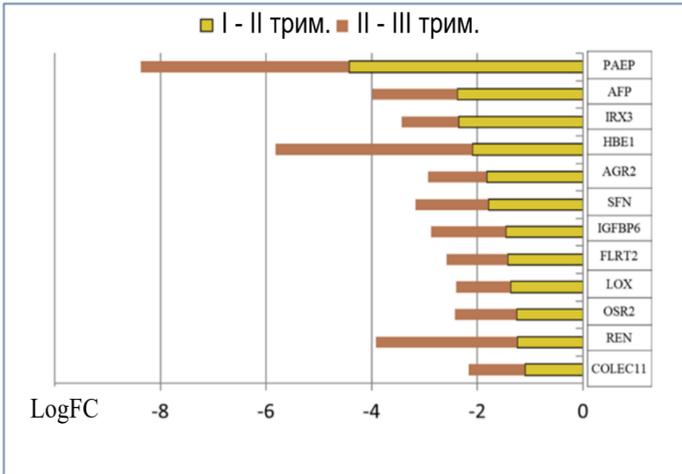


Рис. 17.
Експресія диференційно експресованих генів з групи DownDown у першій і другій половині вагітності

DownUp (рис. 18). Експресія трьох генів спершу знижена у 1_2, але зростає у періоді 2_3. Лептин, муцин 20 і тканинний активатор плазміногену координують енергетичний баланс, бар'єрні властивості трофобласта та контроль згортання крові, готуючи комплекс «мати – плацента – плід» до пологів.

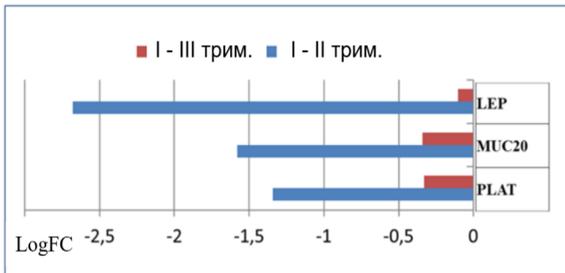


Рис. 18. Експресія диференційно експресованих генів з групи DownUp у першій і другій половині вагітності

Аналіз і узагальнення результатів

Створена й регулярно оновлювана база IGEA, доповнена матеріалами з публікацій, листування з авторами та даними про стать плоду, дає змогу одразу знаходити зразки матково-плацентарного комплексу з потрібними характеристиками, оминаючи трудомістку ручну стандартизацію, необхідну в GEO чи ArrayExpress. Саме на використанні зразків з бази IGEA базується це дослідження. Базу можна використовувати для вторинного аналізу генної експресії інших компонентів матково-плацентарного комплексу. Нині ресурс

охоплює лише мікроарейні дані й прееклампсію; наступний крок — додати RNA-seq та інші ускладнення вагітності.

Порівняльний аналіз експресії генів між інтервалами 1_2 та 2_3. Інтегративний аналіз мікроарей-даних показав, що у плаценті складним чином переплітаються чотири головні програми ко-експресії:

- **Морфогенез і комітування.** У періоді 1_2 активуються каскади Wnt, Hippo, Notch, BMP та гомеобокс-фактори транскрипції, які визначають клітинну структуру і формують тривимірну архітектуру ворсин і судин. У 2_3 їхня експресія різко спадає, що відбиває завершення формування клітинної і просторової структури плаценти.
- **Проліферація й запалення.** Прозапальні хемокіни та цитокіни («UpDown») збільшують експресію у періоді 1_2, коли йде активне формування і розгалуження плацентарних судин, прозапальні процеси із залученням імунних клітин та збільшення маси органу (ріст). В інтервалі 2_3 експресія цих генів йде на спад.
- **Фактори росту й ендокринна зрілість.** Гени «UpUp» безупинно нарощують активність обома інтервалами й забезпечують зрілу синцитіальну функцію (антибактеріальний бар'єр, згортання крові, ліпідний обмін та гормональний статус).
- **Згасання ранніх програм.** Кластери «DownDown» (*PAEP*, *AFP*, *HBE1* тощо) затухають від самого початку, що свідчить про відпрацьовану роль імплантаційних механізмів і ранньої плацентарної активності, яку пізніше перебирає на себе плод. Натомість «DownUp» (*LEP*, *MUC20*, *PLAT*) вдруге активуються у 2_3, готуючи тканину до пологів.

Переломний момент між 1_2 і 2_3. Разюча зміна траєкторії генної експресії припадає саме на межу 1_2/2_3 – час, коли з'являються перші клінічні прояви плацентарної недостатності та прееклампсії. Одна з ймовірних причин полягає у дисбалансі потреб плода й ресурсів плаценти (рис. 19).

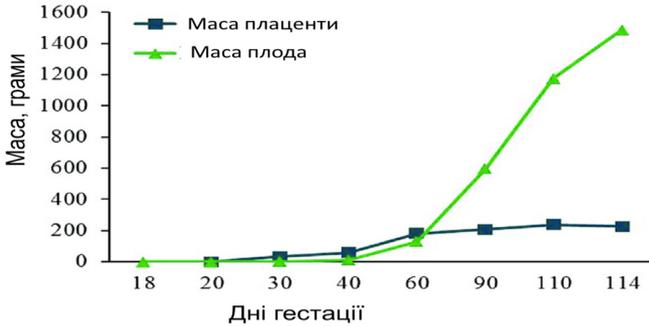


Рис. 19. Інтенсивність набору маси плодом (прямокутники) і плацентою (трикутники). Після середини гестації маса плода збільшується швидше, ніж маса плаценти й ця «ножицеподібна» різниця збільшується до кінця вагітності. Такий розрив посилює чутливість системи до будь-якої дисфункції плаценти і є одним з пояснень, чому клінічні ознаки виявляються саме після 20 тижнів (*Knox et. al., 2008*).

Ключовий висновок. Перебудова плацентарного транскриптома від «будівельної» програми 1_2 до «функціональної» 2_3 проходить через чітку переломну фазу, синхронізовану з різким зростанням потреб плода; саме цей момент є критичним в патогенезі пізніх ускладнень вагітності.

Диференційно експресовані гени, асоційовані з прееклампсією. Порівняння 253 ДЕГів інтервалу 1_2 і 489 ДЕГів інтервалу 2_3 з 376 генами, асоційованими з прееклампсією, дало відповідно 23 та 38 генів у перетині. Така велика частка спільних генів (понад 9 % у 1_2 і 8 % у 2_3) пояснюється тим, що саме ці молекули у нормі керують ключовими плацентарними процесами: тонусом судин (*REN*), згортанням крові (*SERPINB2*), запаленням (*IL6*) та хемотаксисом. Якщо їхній природний ритм експресії збивається, вони автоматично переходять із «фізіологічного» списку в «патологічний» набір.

Асоційовані з прееклампсією ДЕГи 1_2 групуються навколо сигнального шляху IL-17 (переважно ранній прозапальний профіль експресії), тоді як ДЕГи 2_3 — довкола IL-10, що відповідає за пізніший протизапальний стан імунної толерантності. Значна кількість цих генів секретується. Через них плацента взаємодіє з організмом матері, а в разі прееклампсії призводить до генералізованого ураження організму матері.

Отже, великий перетин не є випадковістю: плацента користується тим самим набором регуляторів для нормального розвитку, який, за певних умов, починає сигналізувати про патологію.

ВИСНОВКИ

Створено унікальну спеціалізовану базу метаданих для вторинного аналізу транскриптому у зразках матково-плацентарного комплексу та крові матері й дитини. Проведено системний інтегративний аналіз генної експресії в плаценті, який виявив загальні зміни між першим і другим та другим і третім триместрами нормальної вагітності, а також закономірності ко-експресії генів впродовж вагітності та взаємодії генів у межах одного біологічного процесу.

1. Створено спеціалізовану базу стандартизованих метаданих (IGEA) з 48 датасетів, яка налічує 1466 зразків для зразків матково-плацентарного комплексу та крові матері й дитини, яка уможливорює простий та швидкий відбір даних за обраними користувачем характеристиками зразків.
2. На підставі даних експресії генів, що були відібрані з обраних з бази IGEA датасетів восьми досліджень цілісної тканини плаценти першого, другого та третього триместрів, було утворено новий інтегрований датасет. За допомогою аналізу диференційної експресії в інтегрованому датасеті отримано оригінальні дані, що кількісно характеризують зміни в експресії 253 генів між першим і другим та 489 генів між другим і третім триместрами вагітності. Встановлено, що, перший часовий інтервал характеризується активними імунними і захисними реакціями, диференціюванням клітин і ростом органу, прозапальними процесами і підтриманням взаємної толерантності у ділянці зіткненням плодової частини плаценти з децидуальною тканиною матері. У свою чергу, другому інтервалу притаманне суттєве зменшення прозапальних процесів, на заміну яким приходять протизапальні процеси, зменшення проліферативної активності й активності морфогенів, генів комітування та активація метаболічних процесів, характерних для високо-диференційованої тканини.
3. Завдяки порівняльному аналізу диференційно експресованих генів між першим і другим та другим і третім триместрами встановлено 54 гени, які змінюють значення експресії на обох інтервалах, і вперше виокремлено чотири типи ко-експресії генів впродовж вагітності: експресія генів неухильно зростає впродовж усього періоду вагітності (18 генів), неухильно знижується від першого до другого і від другого до третього триместру (12 генів) і гени, траєкторія експресії яких змінюється всередині вагітності від збільшення у першій половині до зменшення у другій половині вагітності (21 ген) і навпаки (3 гена).
4. Вперше виокремлено типи взаємодії диференційно експресованих генів, які функціонують в межах одного біологічного процесу: взаємне виключення (активація прозапальних хемокінів (11 генів) супроводжується зниженням експресії гомеостатичних хемокінів (4 гени) у першій половині вагітності) і синергія (активація активаторів

- морфогенезу (12 генів) з одночасним інгібуванням їх інгібіторів (9 генів) у першій половині вагітності або активація субодиниць інгібіну (2 гени) при інгібуванні активіну.
5. Вперше впродовж розвитку плаценти встановлено зміну експресії генів комітування, які є активними в першій половині (6 генів) та знижують експресію у другій половині вагітності (5 генів).
 6. Вперше серед диференційно експресованих генів між першим і другим та другим і третім триместрами нормальної вагітності визначено гени, які асоційовані з преєклампсією, і які з них кодують білки, що секретуються в кров.
 7. Аналіз диференційної експресії генів в отриманому в даному дослідженні інтегрованого датасету продемонстрував зниження експресії плейотропного маркерного гена лептину (*LEP*) у першій половині вагітності, яке вважалося суперечливим.

ПЕРЕЛІК НАУКОВИХ ПРАЦЬ ВИКОНАНИХ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Статті у наукових фахових виданнях

1. **Lykhenko, O.**, Frolova, A., & Obolenskaya, M. (2021) ‘Changes in the human placental transcriptome during the physiological course of pregnancy’, *Biopolym. Cell*, 37(1), pp. 73–82. Available at: doi:10.7124/bc.000A4D. (Scopus, Q4) *(Особистий внесок здобувача – застосування розробленого пайплайну до даних експресії в здоровій плаценті першого, другого та третього триместрів, біологічна характеристика знайдених диференційно експресованих генів, написання статті та підготування її до друку).*
2. **Lykhenko, O.**, & Obolenskaya, M. Yu. (2021) ‘The knowledge of chromosomal sex is important for large-scale analysis of gene expression’, *Dopov. Nac. akad. nauk Ukr.*, 2021(1), pp. 100–109. Available at: doi:10.15407/dopovidi2021.01.100. *(Особистий внесок здобувача – визначення статі плоду за даними експресії, покращення якості пайплайну інтегративного аналізу завдяки даним про стать, написання статті та підготування її до друку).*
3. **Lykhenko, O.**, Frolova, A., & Obolenskaya, M. (2021) ‘Consecutive integration of available microarray data for analysis of differential gene expression in human placenta’, *Biotechnologia Acta*, 14(1), pp. 38–45. Available at: doi:10.15407/biotech14.01.38. *(Особистий внесок здобувача – програмування пайплайну обробки та інтеграції даних експресії, написання статті та підготування її до друку).*
4. **Lykhenko, O.**, Frolova, A., & Obolenska, M. (2017) ‘Designing the database for microarray experiments metadata’, *2017 IEEE International Young Scientists Forum on Applied Physics and Engineering (YSF)*. IEEE. Available at: doi:10.1109/YSF.2017.8126658. *(Особистий внесок здобувача – розробка бази метаданих IGEA, написання статті та підготування її до друку).*
5. **Lykhenko, O.**, Frolova, A., & Obolenskaya, M. (2017) ‘Creation of gene expression database on preeclampsia-affected human placenta’, *Biopolym. Cell*, 33(6), pp. 442–452. Available at: doi:10.7124/bc.000967. (Scopus, Q4) *(Особистий внесок здобувача – відбір датасетів для дослідження, стандартизація та розширення метаданих, написання статті та підготування її до друку).*

Тези наукових конференцій

1. Frolova A, **Lykhenko O.**, Bondarenko V., and Obolenskaya M. (2017) ‘Step by step to integration of gene expression profiles for preeclampsia-affected human placenta’, *China — Ukraine International Symposium on Innovation and Technology*, 30.10 — 03.11. 2017. Shandong Academy of sciences. Abstract book. pp. 1–8.
2. Зубенко, С.В., **Лихенко, О.К.**, Фролова, А.О. та Оболенська, М.Ю. (2021) ‘Comparing alternative pipelines for RNA-seq differential expression analysis between preeclamptic and control placental samples’. У: XV Всеукраїнська конференція молодих вчених з міжнародною участю, присвячена 30-

річчю Незалежності України, 26 травня 2021, ІМБіГ НАН України, Київ.

3. **Lykhenko, O.**, Zhoha V., Lykhenko DK., Frolova A., Martsenyuk O., Obolenska M. (2019) «How smoking may decrease the risk of preeclampsia development». *Medical and Clinical Chemistry Scientific Journal*, I. Horbachevsky Ternopil National Medical University Institute of Pharmacology and Toxicology of NAMS of Ukraine, 3(80), vol. 21, p. 110.
4. Жалілова, Є.О., **Лихенко, О.К.** та Оболенська, М.Ю. (2023) ‘Топ-двадцять диференційно експресованих генів у плаценті людини між другим і першим триместрами фізіологічної вагітності’. У: «Біотехнологія XXI століття»: матеріали XVII Міжнародної науково-практичної конференції, 19 травня 2023, Київ: КПІ ім. Ігоря Сікорського, секція 2, с. 225–227.
5. Кукуруза, Є.О., **Лихенко, О.К.** та Оболенська, М.Ю. (2023) ‘Алгоритм пошуку тканинно-специфічних генів, продукти яких секретуються у кров, на прикладі плаценти людини’. У: «Біотехнологія XXI століття»: матеріали XVII Міжнародної науково-практичної конференції, 19 травня 2023, Київ: КПІ ім. Ігоря Сікорського, секція 2, с. 228–232.
6. Жалілова, Є.О., **Лихенко, О.К.** та Оболенська, М.Ю. (2024) ‘Pathway enrichment analysis in the human placenta for the period between the second and first trimesters of physiological pregnancy’. Тези Міжнародної конференції з нейронаук та наукових читань, присвячених вісцеральній фізіології та патофізіології, 19–21 листопада 2024, Київ. *Фізіол. журн.*, 70(5) дод., с. 33.
7. Martynenko, V.V., Zhalilova, Y.O., **Lyhenko, O.K.** & Obolenskaya, M.Yu. (2024) ‘Top 20 differentially expressed genes across trimesters of normal pregnancy’, in *Proceedings of the International Scientific Conference BioGENext*, Kyiv, Ukraine, 17–20 September 2024. *Biopolymers and Cell*, 40(3), p. 210. doi: 10.7124/bc.000AEA.

АНОТАЦІЯ

Лихенко О.К. Спеціалізована база даних з генної експресії у матково-плацентарному комплексі та крові матері й дитини та плацентарний транскриптом упродовж нормального перебігу вагітності. — Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата біологічних наук за спеціальністю 03.00.03 (молекулярна біологія). — Інститут молекулярної біології і генетики НАН України, Київ, 2025.

Роботу присвячено створенню спеціалізованої бази даних шляхом збору наявних даних генної експресії у зразках матково-плацентарного комплексу та крові матері й плода та інтегративному аналізу транскриптому плаценти упродовж нормального перебігу вагітності.

На підставі використаних даних генної експресії, що містяться у репозиторіях Gene Expression Omnibus (GEO) та ArrayExpress, було виконано аналіз консолідованих даних, що дозволило збільшити об'єм вибірки і поліпшити статистичну вагомість результатів аналізу. Цей аналіз первинних даних генної експресії з декількох досліджень (інтегративний аналіз) став можливим завдяки відбору, стандартизації (за онтологіями MeSH і EFO) та розширенню метаданих для зразків в матково-плацентарному комплексі та крові матері й плоду, для яких існують зразки з описаною генною експресією мікрмасивів. Для подальшого вільного доступу до стандартизованих метаданих за допомогою програмних засобів Python/Django/Postgres розроблено реляційну базу метаданих даних IGEA та веб-веб сайт для доступу для неї. База уможливує простий та швидкий відбір датасетів за обраними користувачем характеристиками зразків.

На підставі даних експресії генів, що були відібрані з обраних з бази IGEA датасетів восьми досліджень цілісної тканини плаценти першого, другого та третього триместрів, було утворено новий інтегрований датасет. За допомогою аналізу диференційної експресії в інтегрованому датасеті отримано оригінальні дані, що кількісно характеризують зміни в експресії 253 генів між першим і другим та 489 генів між другим і третім триместрами вагітності. Встановлено, що, перший часовий інтервал характеризується активними імунними і захисними реакціями, диференціюванням клітин і ростом органу, прозапальними процесами і підтриманням взаємної толерантності у ділянці зіткненням плодової частини плаценти з децидуальною тканиною матері. У свою чергу, другому інтервалу притаманне суттєве зменшення прозапальних процесів, на заміну яким приходять протизапальні процеси, зменшення проліферативної активності й активності морфогенів, генів комітування та активація метаболічних процесів, характерних для високодиференційованої тканини.

Показано й уперше кількісно охарактеризовано, що у другому триместрі відбувається перелом у експресії генів 24 генів. Напрямок експресії цих генів від першого до другого триместру змінюється на протилежний від другого до третього триместру. Так, на прикладі поліфункціонального гену лептину (*LEP*), було показано зниз

ення рівня експресії у період від першого до другого триместру, із наступним його зростанням від другого до третього триместру. Переломний момент в експресії генів збігається в часі з першим проявом клінічних симптомів. Автор не виключає залежності клінічних симптомів від дисфункції генів, напрямок експресії яких різний у першій і другій половині вагітності, адже серед них, наприклад, є статистично значуща частка генів (в тому числі *LEP*, *IL6*, *CXCL8*, *HSDB2*), порушення експресії яких асоціюють з преєклампсією, поширеним ускладненням вагітності.

За результатами дослідження вперше описано чотири універсальні траєкторії ко-експресії генів («Up-Up», «Up-Down», «Down-Down», «Down-Up») та три характерні схеми взаємодії ДЕГів у межах біологічного процесу: взаємне виключення (активація прозапальних хемокінів проти зниження експресії гомеостатичних, тощо), а також – синергія: активація активаторів морфогенезу з одночасним інгібуванням їх інгібіторів у першій половині вагітності або активація інгібіторів при інгібуванні активінів у другій половині вагітності.

Ключові слова: плацента, транскриптом, гени, інтегративний аналіз, перший триместр, другий триместр, третій триместр, база даних, молекулярна біологія, метаболізм, імунна регуляція, ангиогенез, гени комітування, морфогени, хемокіни.

SUMMARY

Lykhenko O.K. A specialized database of gene expression in the uterine-placental complex, maternal and fetal blood, and the placental transcriptome throughout physiological pregnancy. — Manuscript.

Dissertation for obtaining the degree of Candidate of Biological Sciences, specialty 03.00.03 — Molecular Biology. — Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, Kyiv.

The study is devoted to the creation of a specialized database of gene expression in samples of the uteroplacental complex and maternal and fetal blood and integrative analysis of the placenta transcriptome during physiological pregnancy.

Open gene expression data accumulated in the Gene Expression Omnibus (GEO) and ArrayExpress repositories were used to conduct a secondary analysis of the combined data of the same type, which increases the sample size and statistical significance of the analysis results. The metadata of the gene expression datasets in the uteroplacental complex and maternal and fetal blood using microarray technology

were selected, standardized, and expanded. A specialized relational database of IGEA metadata was created for the first time. The database enables quick and straightforward selection of datasets according to the characteristics of the samples selected by the user.

Gene expression data from the IGEA database, comprising eight studies of bulk placenta tissue across the first, second, and third trimesters, were combined to form a fundamentally new integrated dataset. For the first time, changes in the expression of 253 genes between the first and second trimesters and 489 genes between the second and third trimesters of pregnancy have been quantitatively characterized. Active immune and protective reactions, cell differentiation and organ growth, pro-inflammatory processes, and maintenance of mutual tolerance in the area of contact of the fetal part of the placenta with the decidual tissue of the mother characterize the first time interval. The second time interval is characterized by a significant decrease in pro-inflammatory processes, a decrease in proliferative activity and the activity of morphogens and commitment genes, and the activation of metabolic processes characteristic of highly differentiated tissue. In the second trimester, a turning point in the expression of numerous genes has been shown and quantitatively characterized for the first time; the direction of gene expression from the first to the second trimester changes to the opposite direction from the second to the third trimester. Among such genes is, for example, the polyfunctional leptin gene (LEP), the expression of which decreases from the first to the second trimester and increases from the second to the third trimester. The turning point in gene expression coincides in time with the first manifestation of clinical symptoms, which does not exclude the dependence of clinical symptoms on the dysfunction of genes, the direction of expression of which is different in the first and second halves of pregnancy.

For the first time, four universal trajectories of gene co-expression (“Up-Up”, “Up-Down”, “Down-Down”, “Down-Up”) and three characteristic patterns of DEG interaction within a biological process have been described: mutual exclusion (for example, pro-inflammatory chemokines versus homeostatic ones), synergy (activation of morphogenesis activators with simultaneous inhibition of their antagonists, and activation of inhibins leads to specific inhibition of activins and suppression of morphogenetic processes).

Keywords: placenta, transcriptome, genes, integrative analysis, database, molecular biology, metabolism, immune regulation, angiogenesis, commitment genes, morphogens, chemokines.

Підписано до друку 25.08.2025 р. Формат 60x90 /16
Ум. друк. арк.0,9. Обл-вид. арк 0,9
Наклад 100 прим. Замовлення № 112-2508817
Віддруковано в поліграфії "КПІ-принт"
03056, м. Київ, просп. Берестейський, 31
Тел.: +38 (068) 308 88 75
<https://kpi-print.com.ua>. E-mail: kpi@kpi-print.com.ua

