

великими проблемами для економіки нашої країни. Таким чином, для забезпечення добробуту громадян нашої країни важливо вивчати механізми старіння та потенційне лікування з метою продовження не лише тривалості життя, а й періоду життя без хвороб нашого власного народу.

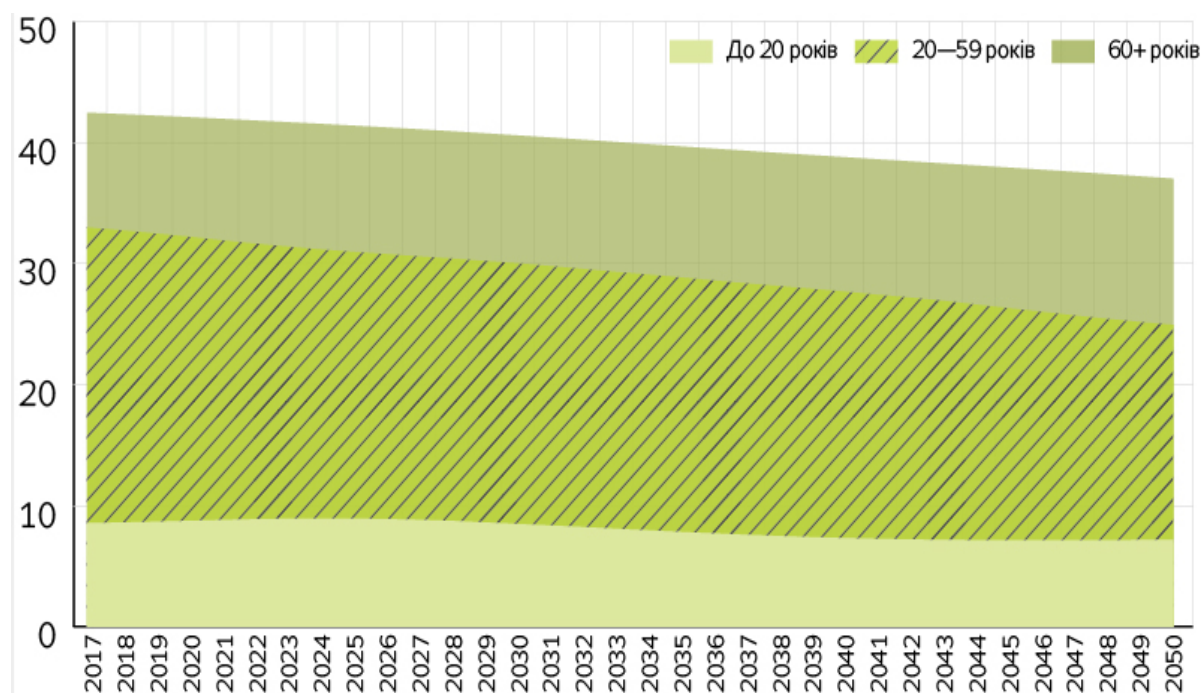


Рис. 1.1. Діаграма з областями, що зображає прогноз динаміки населення України різних вікових груп на період до 2050 р., млн чол.: світло-зеленим позначено вікову групу до 20 років, зеленим із штриховкою – 20-59 років, темно-зеленим – 60 та більше років (за даними “Комплексного демографічного прогнозу України до 2050 р.”, 2006 р.)

1.2 Теорії старіння

За часи людства було запропоновано багато теорій старіння. Чудова спроба класифікувати запропоновані теорії була зроблена в 1990 році та перерахувала близько 300 різних теорій (Medvedev, 1990). Однак немає єдиної теорії, що могла б повністю пояснити старіння. Тому на сьогодні вважається, що усі вони доповнюють один одного, а не є взаємовиключними, оскільки всі

РОЗДІЛ 2

МАТЕРІАЛИ ТА МЕТОДИ ДОСЛІДЖЕННЯ

2.1 Об'єкт дослідження

Об'єктом даного дослідження є одні із найбільших загальнодоступних даних експресії генів з віком, що налічують 269 зразків префронтальної кори головного мозку людини та були отримані із використанням технології мікрочіпа (Colantuoni, 2011). Зразки було зібрано лише у здорових людей та вони охоплюють весь період життя людини, починаючи від ембріонального розвитку (пренатальні зразки) до старіння (80 років). У поточному аналізі ми навмисно виключили дані, отримані на стадії ембріонального розвитку, дитинства та раннього дорослого віку до 20 років, що б виключити з аналізу зміну у експресії, що опосередкована процесами розвитку головного мозку, які тривають до початку дорослого віку (близько 20 років).

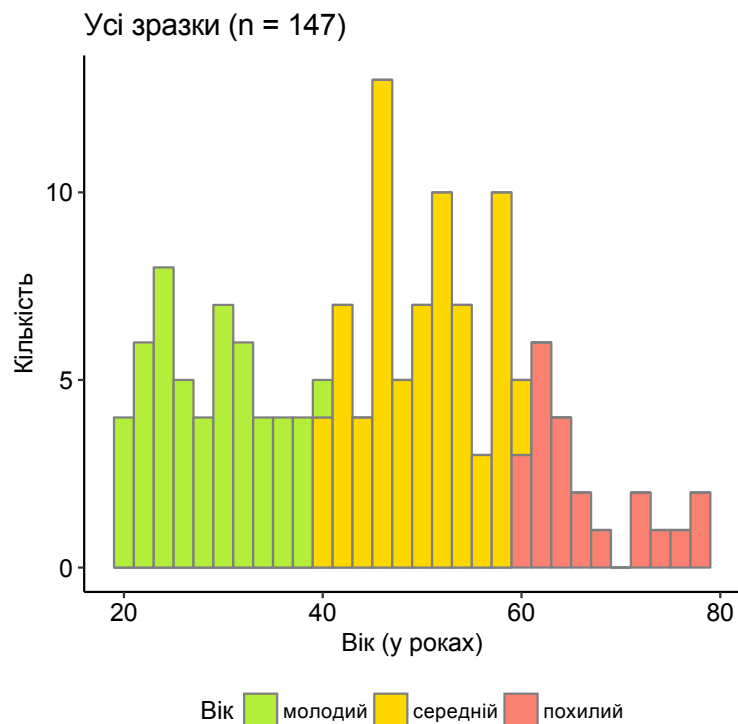


Рис. 2.1. Гістограма, що відображає розподіл зразків за віком в наборі даних транскриптому префронтальної кори мозку людей різного віку, колір

Надалі, я би хотіла вказати на варіабельність у біологічних та технічних факторах у наборі даних. По-перше, зразки не розподілені рівномірно протягом різних вікових груп. Ми спостерігаємо найвищу щільність зразків у молодому та середньому віці та найнижчу щільність у похилому віці. Далі, зразки обробляли в 19 партіях. Частина зразків у віці від 20 до 80 містить мінімум 2 зразка на партію та максимум 13 зразків / партії. Зразки походять від людей обох статей, чоловічої та жіночої та 4-х різних рас афроамериканської (AA), європеїдної (CAUC), латиноамериканської (HISP) та азійської (AC). Більшість зразків було отримано від людей афроамериканської та європеїдної рас, відповідно 82 і 57 і лише 4 зразка були отримані від латиноамериканської та азійської раси. Нарешті, зразки мали різні вимірювання якості, а саме pH та РНК цілісність (RNA integrity number (RIN)), і відрізнялися часом збирання після смерті (post-mortem interval (PMI)).

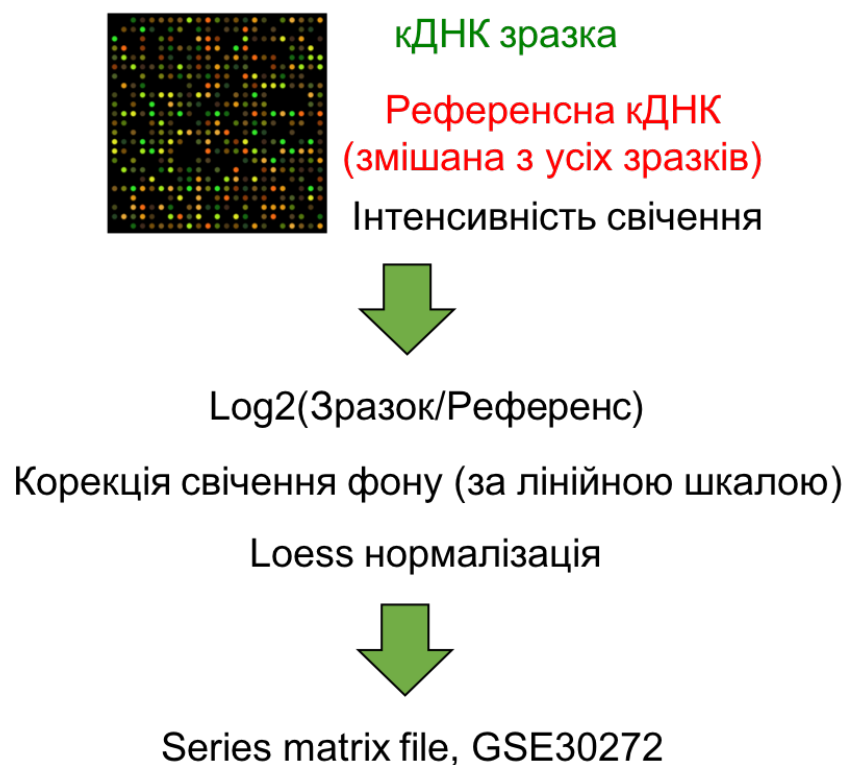


Рис. 2.1. Схема попередньої обробки даних транскриптому мозку людини до отримання Series Matrix File (Colantuoni, 2011).

Для того, щоб вирішити, чи потрібно корегувати варіабельність даних, що зумовлена вищенаведеними біологічними та технічними факторами (вона може спотворювати, «маскувати» або навпаки штучно створювати, віковий ефект) ми побудували графіки залежності даних факторів від віку.

На додаток до візуального аналізу даних графіків ми здійснювали тест ANOVA на відмінність віку людей, що належать до різних груп (для категоричних факторів: стать, раса, партія зразків) та (або) рахували коефіцієнт кореляції Спірмена між віком та відповідною змінною (для чисельних факторів: RIN, pH, pmi).

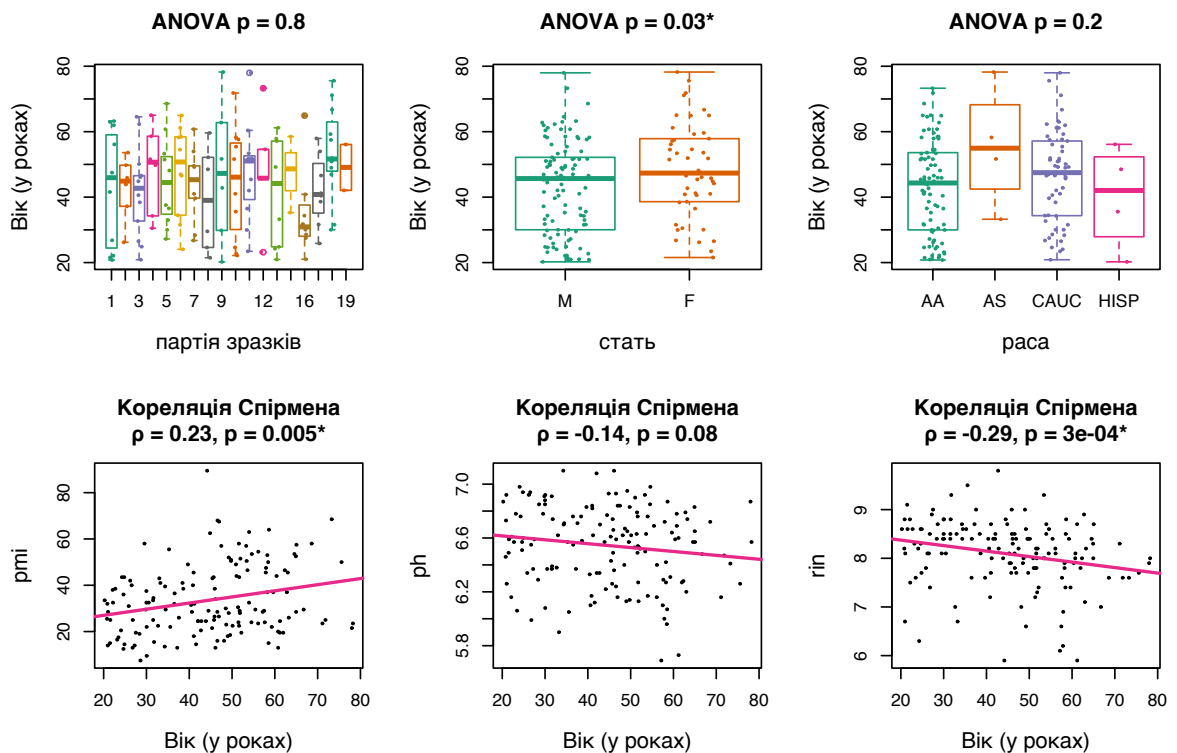


Рис.2.1. Графіки, що зображають залежність між віком та категоріальними та чисельними факторами у даних транскриптому мозку людини: Тест ANOVA проводили, щоб з'ясувати, чи існує різниця у віці між зразками з різних партій, статі та раси; для безперервних змінних, таких як інтервал після смерті (pmi), цілісність РНК (rin) та pH було пораховано коефіцієнт кореляції Спірмена (ρ) від віку до змінної, * $p \leq 0,05$

потенційно може спричинити перенавчання моделі. У моєму випадку SVA оцінює 20 сурогатних змінних у даних експресії генів. Я додала їх (сурогатні змінні) як коваріати до лінійної моделі, що пояснювала експресію генів згідно віку:

$$Y_i \sim \beta_0 + \beta_1 * \text{вік}^{0.25} + \beta_{sv1} * SV1 + \dots + \beta_{SV20} * SV20$$

, де Y_i - це лог-нормалізована експресія гену, $i = 1, \dots, n$, β_0 – перетин з віссю ординат, β_1 – коефіцієнт для ефекту віку, β_{sv} - коефіцієнти для сурогатної змінної.

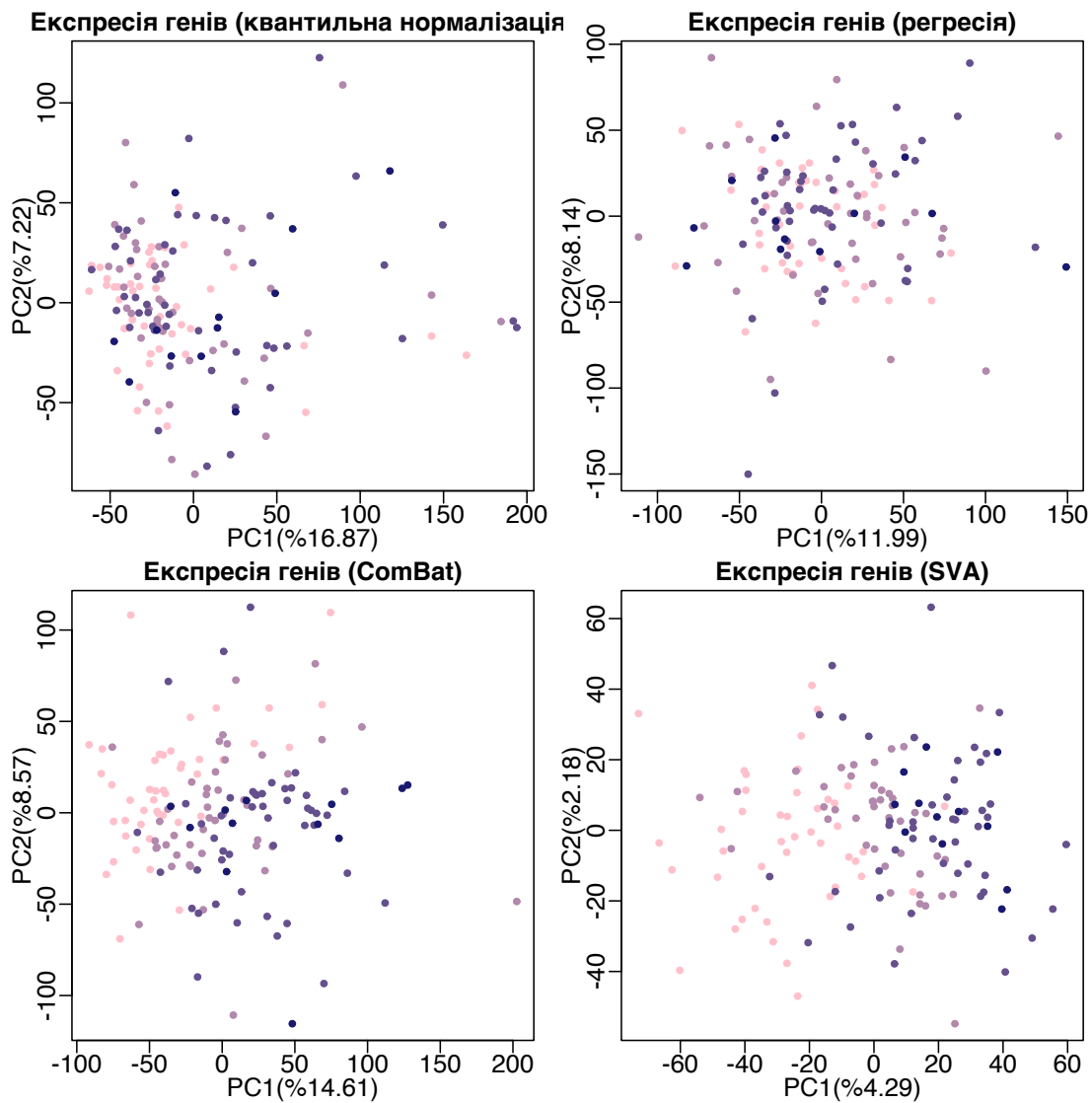


Рис 2.4. Графіки залежності перших двох основних компонент (PC), отриманих після факторного аналізу даних експресії генів мозку людини до і

тестування. Залежно від застосованого методу корекції, лінійна модель також включала різні виміряні чи виведені коваріати (див. Етапи обробки даних) наступної загальної форми:

$$Y_i = \beta_{i0} + \beta_{i1} * \text{вік}^{0.25} + \text{коваріати} + \varepsilon_i,$$

де Y_i - це лог-нормалізована експресія гену, $i = 1, \dots, n$, β_0 – перетин з віссю ординат, β_1 – коефіцієнт для ефекту віку, ε_i – похибка лінійної моделі.

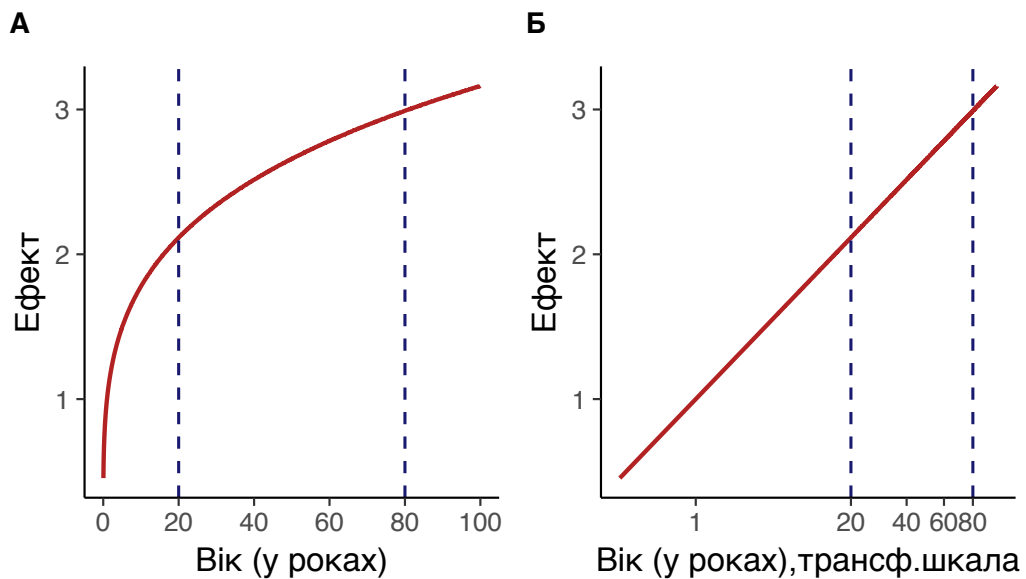


Рис. 2.5. Графіки, що ілюструють залежності $y = x^{0.25}$, яка була використана для диференційного аналізу експресії генів з віком: А - ось абсцис в нормальному масштабі; В - осі абсцис у масштабі $x^{0.25}$; пунктирні лінії обмежують віковий діапазон 20 - 80 років, що використано в даному аналізі.

2.2.5 Диференційна варіабельність

Безперервний підхід: Спочатку ми будували лінійну модель, що намагалась пояснити експресію генів під час старіння, використовуючи вік у ступеню 0.25 та потенційних коваріати. Далі кореляція Спірмена була розрахована між абсолютними значеннями похибок від лінійної моделі та віку. Таким чином, коефіцієнт кореляції Спірмена використовували як міра зміни у

варіабельності експресії генів з віком, $\Delta var(\rho)$. Значення P для кореляційних оцінок Спірмена були виправлені для багаторазового тестування за допомогою FDR. $q \leq 0,05$ використовували як поріг, щоб визначити статистично значимі диференційно варіабельні (ДВ) гени.

$$\Delta var_i(\rho) = \text{оцінка кореляції Спірмена } \rho(|\varepsilon_i|, \text{вік})$$

Груповий підхід: Спочатку, потенційні коваріати були видалені з даних, використовуючи вище наведені методи корекції. Далі було застосовано "Груповий підхід", який ми розробили самостійно. Його метою є порівняння варіабельності експресії генів між літньою (60-80-років) та молодшою (20-40 років) віковими групами із використанням інтерквартильного розмаху (IQR) як міри варіабельності.

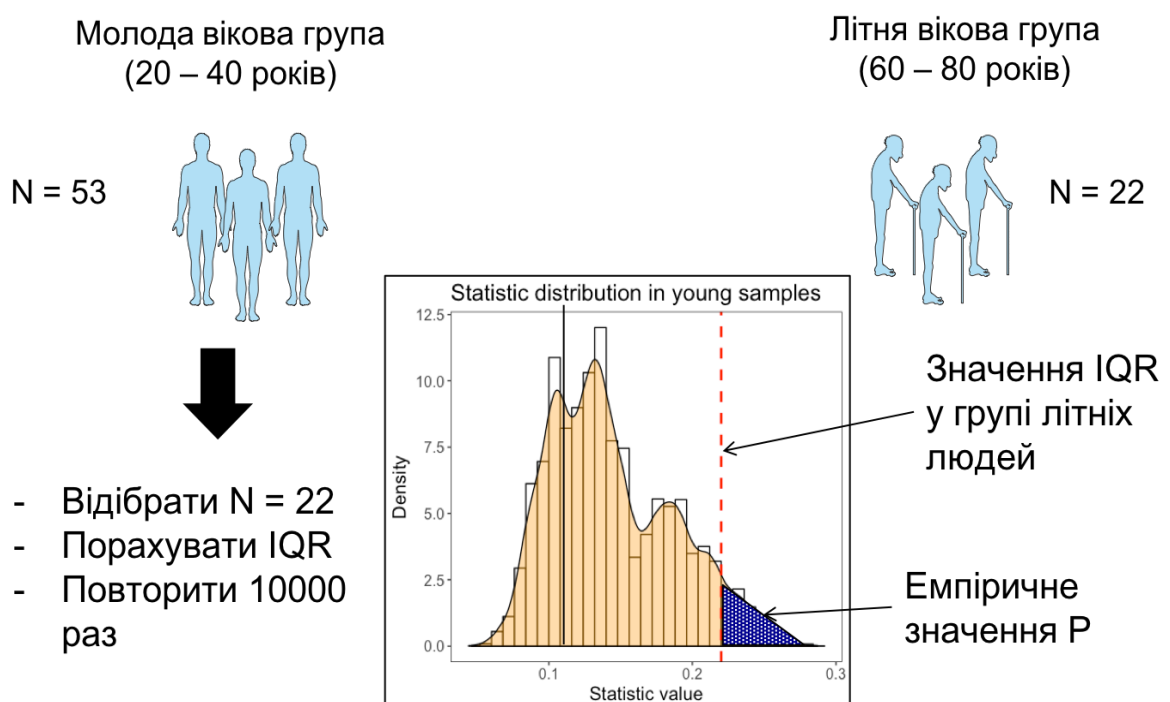


Рис .2.6. Схема «групового підходу» для порівняння варіабельності у експресії генів між молодшою та літньою віковими групами

генів, виявлених після SVA, і 96% генів, виявлених після корекції з використання регресії. Дане перекриття також супроводжується сильною кореляцією між розмірами ефекту ДЕ для регресії та результатами SVA ($\rho = 0,89$, Рис.3.1 Б).

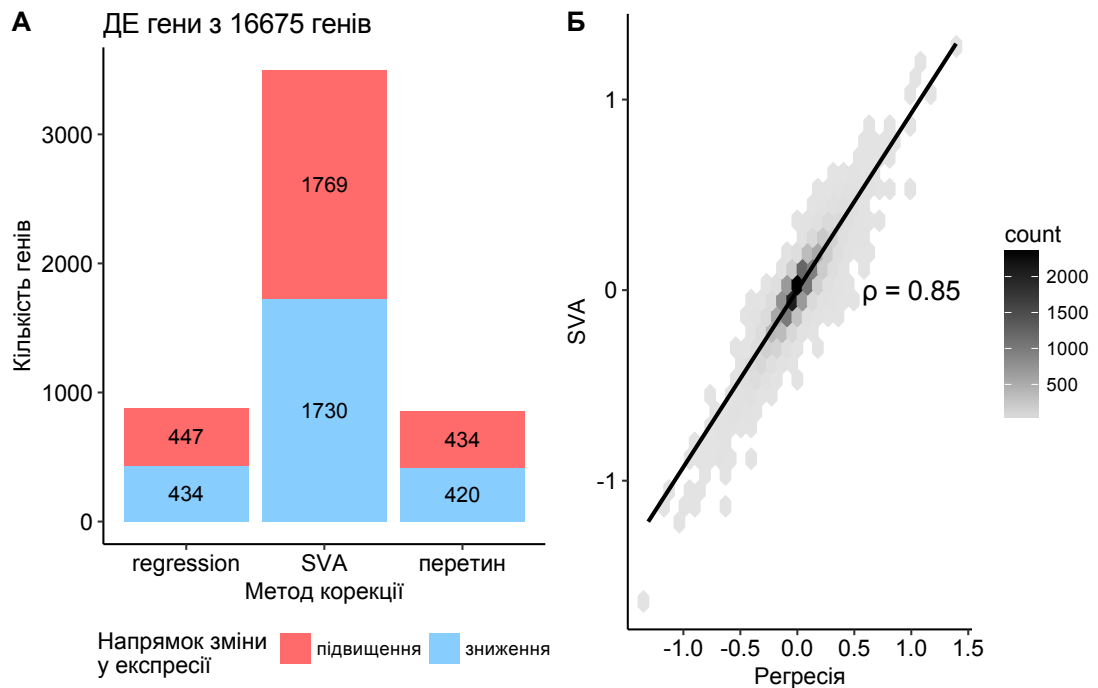


Рис. 3.1. Графіки, що ілюструють диференційну експресію генів у транскриптомі мозку людини із використанням двох методів переобробки (регресія та SVA): А - стовпчаста діаграма диференційно-експресованих генів під час старіння, що були виявлені після застосування попередньої обробки у вигляді регресії, SVA та їх перекриття. Колір відображає напрямок зміни: червоний - гени, що збільшують експресію з віком, сині - гени, що знижують свою експресію з віком, Б – теплокарта, що відображає кореляцію між даними обробленими за допомогою регресії та SVA (дані - величина ефекту з віком для усіх генів (N= 16675))

Проведення аналізу представленості функціональних груп генів за базою даних Генної Онтології (Біологічний процес) для результатів диференційної експресії, отриманих після застосування регресії та SVA (окремо) виявило, що

відбувається пригнічення у експресії 138 функціональних груп генів (є спільними для результатів регресії та SVA). Дані групи (після попереднього скорочення та узагальнення) представлено на Рис. 3.2 у вигляді ієрархічної діаграми TreeMap. Серед представлених генів я хочу звернути увагу на гени, що пов'язані із когнітивними функціями, такими як модуляція синаптичної передачі, навчання та пам'ять, які також переживають пригнічення у експресії і попередньо згадувалися в літературі.

Кількість функціональних груп генів, що підвищують свою експресію з віком є набагато нижчою, порівняно з тими, що пригнічують свою експресію. Також майже відсутнє перекриття між результатами, які отримані на даних оброблених за допомогою регресії та SVA.

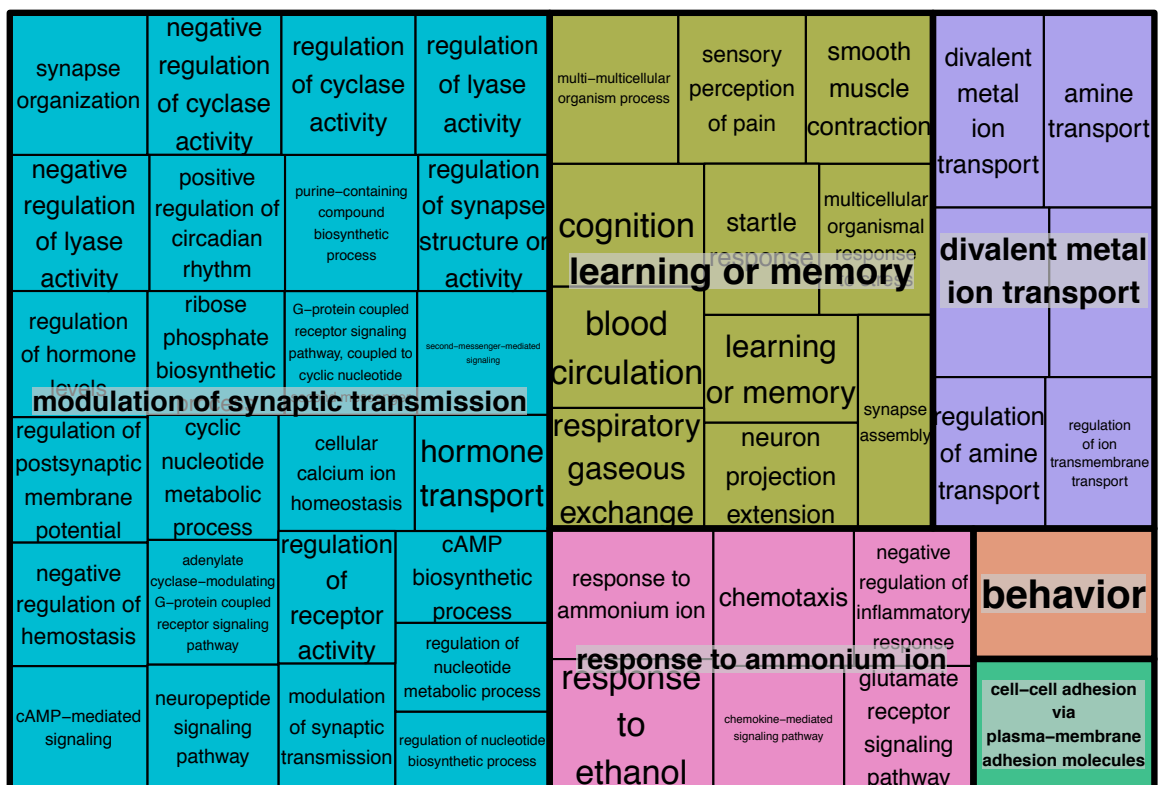


Рис. 3.2. Ієрархічна діаграма TreeMap функціональних груп генів (база даних Gene Ontology, biological process), що пригнічують свою експресію з віком у корі мозку людини (наведено лише терміни GO, які є спільними між регресією та результатами SVA): розмір прямокутників зображує унікальність

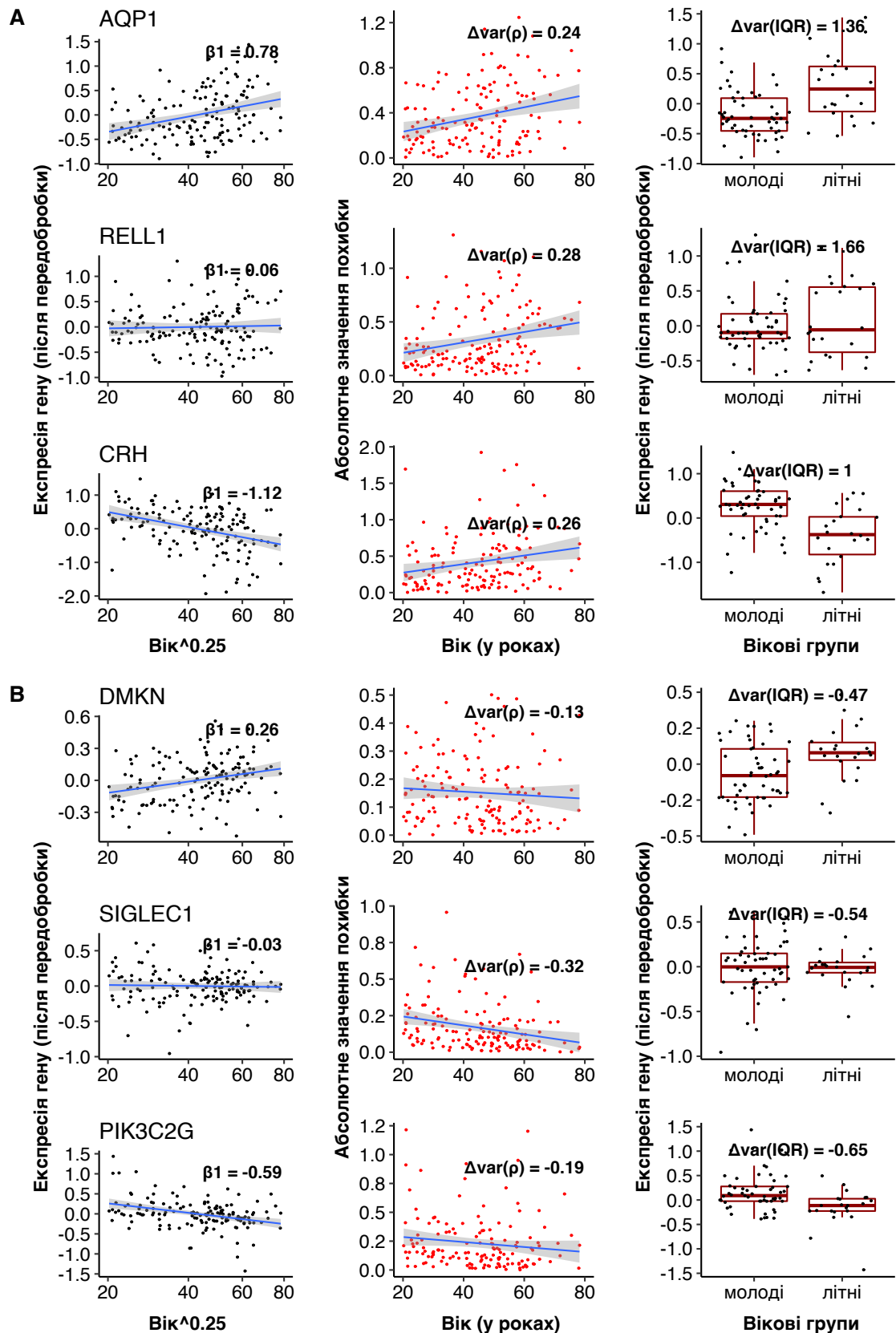


Рис. 3.3 Графіки, що ілюструють диференційну експресію, безперервний та груповий підходи для обчислення диференційної варіабельності з віком на

генів збільшують варіабельність з віком порівняно з тими, що зменшують. Величини зміни у варіабельності з віком для даних, скоректованих за допомогою регресії та SVA, показують слабку кореляцію, ρ ($\Delta\text{var}(\rho)$ - регресія, $\Delta\text{var}(\rho)$ - SVA) = 0,35 (рис. 3.4).

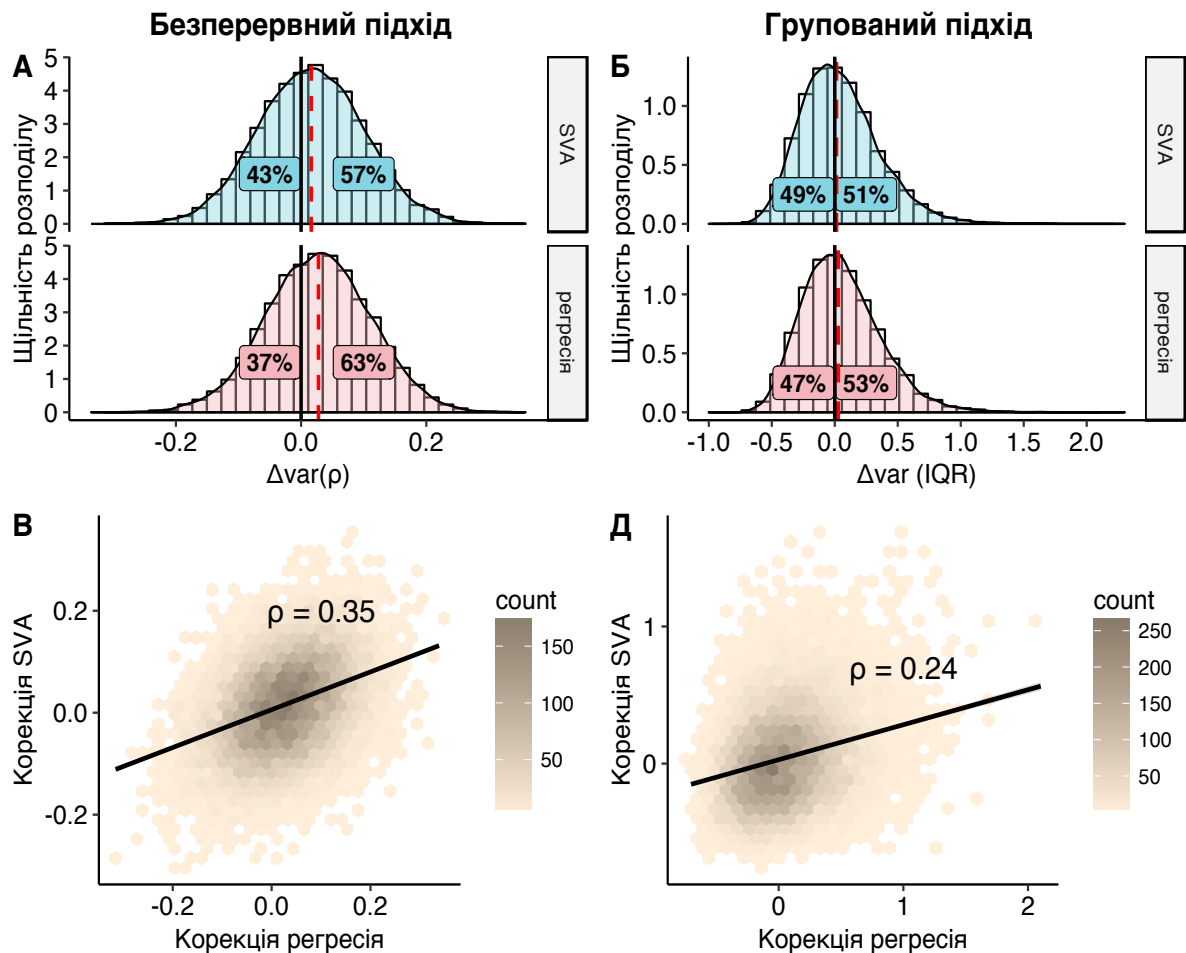


Рис 3.4. Графіки, що ілюструють зміну у варіабельності експресії генів з віком (Δvar) людини із використанням двох методів передобробки даних (регресії та SVA): **А, Б** - розподіл величини зміни у варіабельності експресії гена з віком для безперервного та групового підходів: червоний позначає результати, отримані із попередньою обробкою із використанням регресії, синій - SVA. Чорна пряма лінія вказує на нуль, червона пунктирна лінія - медіана розподілу. **В, Г** - гексагональна теплокарта, що ілюструє зв'язок між Δvar для кожного гена для результатів регресії та SVA; градієнт відображає

3.2.1 Диференційно-варіабельні гени

Далі ми поцікавилися, чи можливо виявити гени, що достовірно змінюють варіабельність своєї експресії з віком. Використовуючи безперервний підхід, ми не виявили генів, що достовірно змінюють варіабельність у експресії з віком після багаторазової корекції тестування. У свій час, груповий підхід виявив 741 та 746 диференційно варіабельних генів (ДВ) генів ($q \leq 0,05$) з використанням попередньої обробки за допомогою регресії та SVA, відповідно (рис. 3.5 А). Проте виявлені набори генів майже не перекриваються: мають лише 83 спільні гени (рис. 3.5 А).

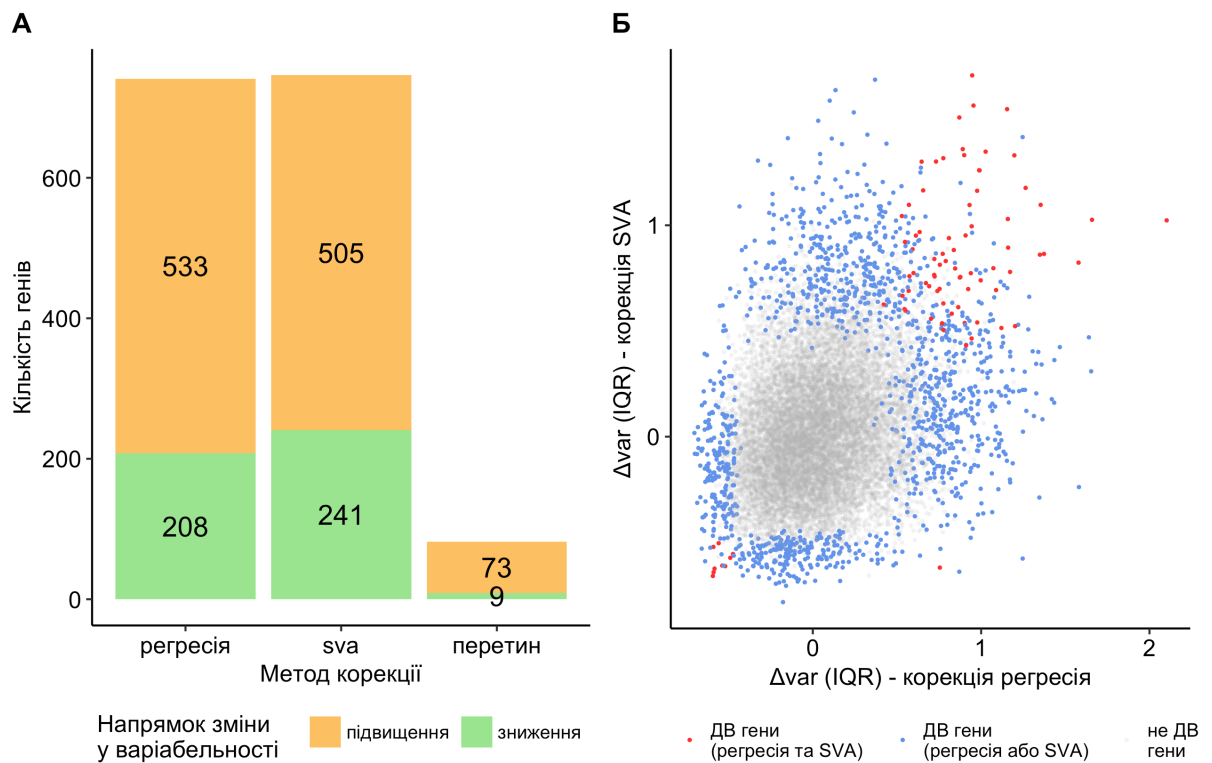


Рис. 3.5. Графіки, що ілюструють статистично значимі зміни у варіабельності експресії гена з віком (Δvar) людини із використанням групового метода обчислення та двох засобів передобробки даних (регресії та SVA): **А** - стовпчаста діаграма, що зображає диференційно варіабельні гени, що були знайдені при обробці даних за допомогою регресії, SVA та їх перетин, напрямок зміни позначено кольором: червоний - збільшення варіабельності,

безперервного (попередня обробка SVA) та групового підходу (попередня обробка за допомогою регресії та SVA).

Однак ми виявили, що гени, які змінюють варіабельність експресії в безперервному підході (попередня обробка за допомогою регресії), частіше зустрічаються у таких шляхах: бактеріальна інвазія епітеліальних клітин, бета-аланіновий обмін, сигнальний шлях Ras, система бактеріальної інвазії епітеліальних клітин ($q < 0,05$). Ці шляхи мають позитивні показники нормалізованого забагачення (NES), що означає, що вони збагачені на гени, які підвищують варіабельність з віком. Крім того, ці шляхи також мають позитивні значення NES у інших підходах, навіть якщо вони не є статистично значущими.

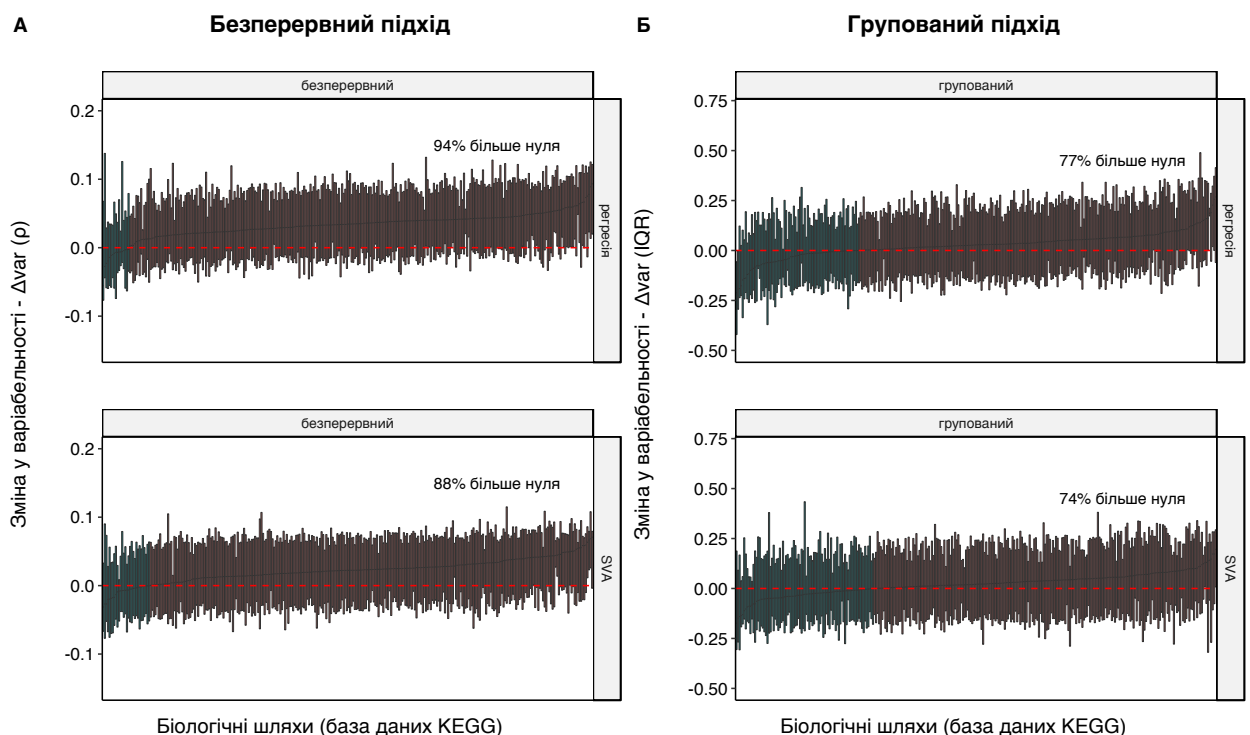


Рис. 3.6 Діаграми, що зображують розподіли величини зміни у варіабельності експресії з віком для генів у біологічних шляхах ($n = 310$), яка порахована за різними методами. Кожна діаграма являє собою коробковий графік, де одна коробка - це розподіл зміни у варіабельності генів для одного шляху. Синій колір позначає шляхи з медіаною нижче нуля, червоний – вище

ДОДАТКИ

Додаток А

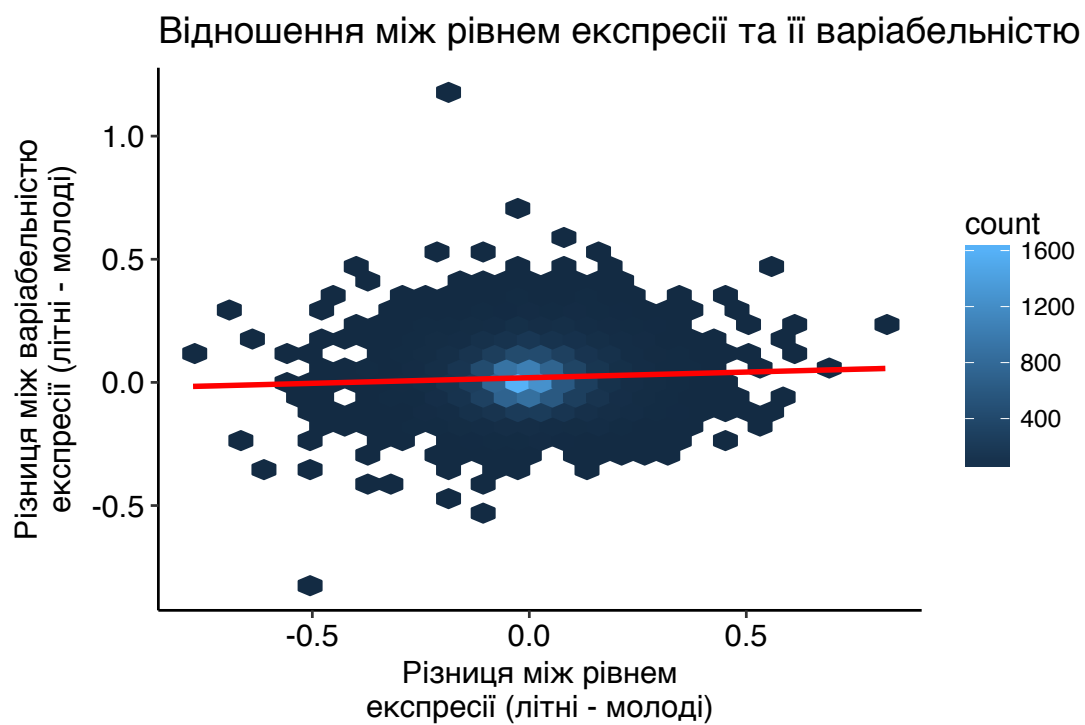


Рис.1. Теплокарта, що показує відношення між різницею у рівні експресії та різницею у варіабельності експресії між літніми та молодими людьми. Градієнт кольору позначає щільність значень.