

ТОП-ДВАДЦЯТЬ ДИФЕРЕНЦІЙНО ЕКСПРЕСОВАНИХ ГЕНІВ У ПЛАЦЕНТІ ЛЮДИНИ МІЖ ДРУГИМ І ПЕРШИМ ТРИМЕСТРАМИ ФІЗІОЛОГІЧНОЇ ВАГІТНОСТІ

Жалілова Є.О.¹, Лихенко О.К.², Оболенська М.Ю.²

¹КПІ ім.Ігоря Сікорського, evgeniy.jalilova@gmail.com

²Інститут молекулярної біології і генетичної інженерії НАН України

Вступ. Плацента – це орган, який розвивається під час вагітності та поєднує організм матері з плодом у триєдину систему – мати – плацента – плод. Порушення функціонування плаценти призводить до різних ускладнень вагітності, таких як затримка росту плоду, передчасні пологи, викидні і смертельно небезпечної для матері і плоду прееклампсії. Перший і другий триместри вагітності є критичними для розвитку дисфункцій плаценти.

Тому з метою виявлення маркерів прееклампсії (найбільш значущої різниці в експресії генів) доцільно проаналізувати диференційно експресовані гени у плаценті людини між першим і другим триместрами фізіологічної вагітності. Для цього було використано метод volcano plot.

Volcano plot – це тип діаграми розсіювання, що об'єднує статистичний показник значущості (p-value), отриманий у результаті статистичного тесту, з величиною кратності зміни (Fold change), який використовують для швидкого виявлення найбільш виражених і найбільш достовірних змін у великих наборах повторюваних даних [1-3]. p-value – ймовірність отримати значення випадкової величини для даної ймовірнісної моделі розподілу. Fold change – це математичне значення, що описує, як кількісно відрізняються дані між собою. Повторювані дані – це дані, які містяться у декількох записах або документах.

Мета роботи – побудувати графік volcano plot і за графіком – визначити диференційно експресовані гени з найбільшим абсолютним значенням lgFC і найменшим значенням p-value для їх подальшого функціонального аналізу.

Матеріали та методи. Дані для побудови volcano plot включають результати порівняльних досліджень експресії генів у плаценті людини між першим та другим триместрами вагітності. Для кожного гена вказаний lgFC (кратність змін) його експресії між другим і першим триместрами і значенням скорегованого p-value з урахуванням поправки на множинну перевірку гіпотез [4]. Розраховували від'ємний логарифм скорегованого p-value у програмі Microsoft Excel. Ці дані були використані для побудови volcano plot.

На основі літературних джерел було розроблено код для програмного забезпечення R [5]. Для побудови volcano plot завантажили і використали пакети з бібліотеки R: ggplot2, readxl, ggrepel: `install.packages("ggplot2", "readxl", "ggrepel")` [5]. Також у ході роботи розраховували π -value для всіх генів з метою відібрати найбільш значущі диференційно експресовані гени: `full $pi.value <- abs(-log10(full $adj.p.val) * full $lgFC)`. π -value – оцінка значущості гена, яка об'єднує FC і p-значення в одну оцінку і пропонує новий спосіб ранжування та вибору генів і позначається грецькою літерою π . π -значення – це добуток від'ємного логарифма скорегованого p-value і lgFC [6].

Результати та обговорення. Точки, що позначені синім і червоним кольорами, являють собою значення, що відображають найбільші зміни у значенні кратності $\lg FC$ (отже, ліворуч або праворуч від центру) і найвищу статистичну значущість (отже, ближче до вершини) (рис.1.).

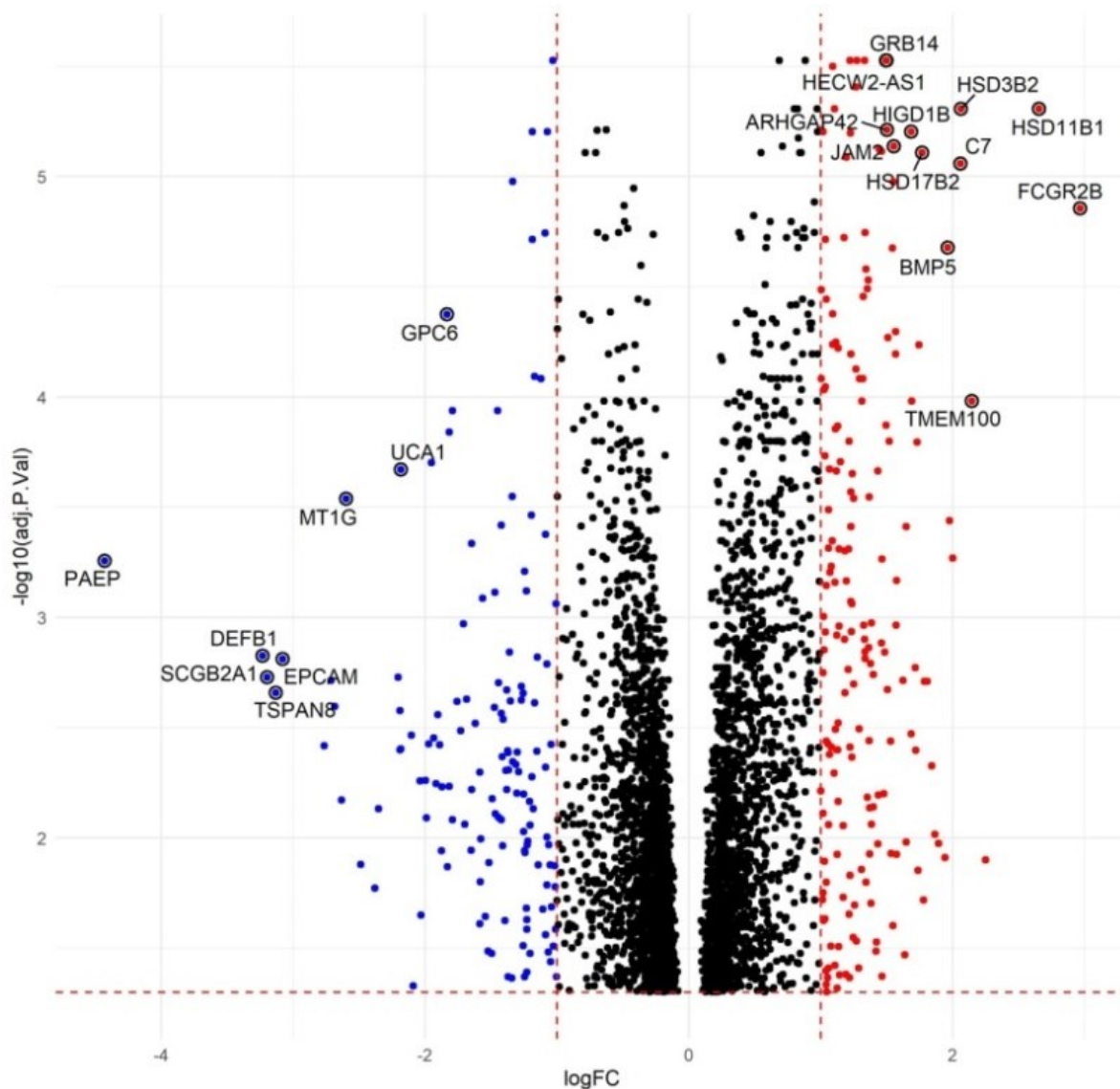


Рис.1 Volcano plot на основі даних про диференційну експресію генів у плаценті людини між першим і другим триместрами вагітності. Вісь x – $\lg FC$, вісь y – $(-\log(\text{скорегованого } p\text{-value}))$; підписані точки – найбільш значущі диференційно експресовані гени.

В результаті побудови volcano plot було виявлено 20 генів з найбільшим абсолютним значенням $\lg FC$ і $-\log(\text{скорегованого } p\text{-value})$, за якими відрізняються плаценти другого і першого триместрів фізіологічної вагітності (табл.1).

Таблиця 1. Найбільш значущі диференційно експресовані гени між 1 та 2 триместрами вагітності. $-\log(\text{adj.p-value}) > 0.05$, $|\log\text{FC}| > 1$

SYMBOL	lgFC	p-value	adj.p-val	$-\log(\text{adj.p-val})$	π -value	1 Trim. Average	2 Trim. Average
FCGR2B	2,9657654	3,0351E-08	1,3976E-05	4,854622245	14,39767081	5,976413251	8,942178677
HSD11B1	2,6555847	3,5506E-09	4,9362E-06	5,306608407	14,09214816	5,146099813	7,801684525
TMEM100	2,1482081	8,3848E-07	0,00010456	3,9806228	8,551206133	5,022407135	7,170615233
HSD3B2	2,0619921	2,8687E-09	4,9362E-06	5,306608407	10,94218475	4,390712316	6,452704443
C7	2,0594369	1,6424E-08	8,757E-06	5,057643463	10,41589781	5,908563815	7,96800076
BMP5	1,9622549	7,226E-08	2,1079E-05	4,676142059	9,175782543	6,359848811	8,322103684
HSD17B2	1,7692687	1,3835E-08	7,7958E-06	5,108137915	9,037668715	8,243213052	10,01248179
HIGD1B	1,6870528	6,9634E-09	6,2549E-06	5,203782584	8,779056038	4,750243269	6,43729608
JAM2	1,5564266	9,7111E-09	7,2865E-06	5,137478944	7,996108842	6,180827669	7,737254261
ARHGAP42	1,5028579	5,4741E-09	6,1617E-06	5,210298846	7,830338818	4,687741411	6,190599317
GRB14	1,5004436	5,5365E-10	2,9841E-06	5,525189393	8,290234855	3,363132425	4,863575988
HECW2-AS1	1,4961308	1,0539E-09	2,9841E-06	5,525189393	8,266405878	2,404174793	3,900305566
GPC6	-1,8322594	1,9436E-07	4,2272E-05	4,373950664	8,014212302	5,774187522	3,941928103
UCA1	-2,180902	2,6322E-06	0,00021285	3,671933787	8,008127813	8,563952473	6,383050453
MT1G	-2,5976813	4,2089E-06	0,00029005	3,537534317	9,189386904	9,481021839	6,883340494
EPCAM	-3,0775269	5,3628E-05	0,00154561	2,810899262	8,650617957	6,933513547	3,855986695
TSPAN8	-3,129142	8,9696E-05	0,0021922	2,659120684	8,320766146	6,144045022	3,014903049
SCGB2A1	-3,1968019	6,9623E-05	0,00186345	2,729682052	8,726252892	5,872264668	2,675462723
DEFB1	-3,2314056	5,03E-05	0,00148986	2,826855792	9,134717707	7,641662289	4,410256664
PAEP	-4,4274681	1,1288E-05	0,00055378	3,256659661	14,41875689	10,75052577	6,323057634

Примітка. lgFC – логарифм кратності змін гена, p-value – ймовірність отримати значення випадкової величини, adj.p-val – скореговане значення p-value з урахуванням поправки на множинну перевірку гіпотез, $-\log(\text{adj.p-val})$ – від’ємний десятковий логарифм скорегованого p-value, π -value – оцінка значущості гена, яка об’єднує FC і p-значення в одну оцінку, 1 та 2 Trim. Average – середні значення логарифмованої матриці експресії генів у першому та другому триместрах.

За допомогою набору інструментів String-db (functional protein association networks) було знайдено асоціації білків та проаналізовано функції генів.

Висновки. 1. Метод volcano plot дав змогу виявити гени з найбільш значущими змінами в експресії у плаценті між другим і першим триместрами вагітності і забезпечив швидкість виявлення і його наочність.

2. Набір інструментів String-db дав змогу проаналізувати функції та взаємозв’язок між білками, що кодуються отриманими генами. Далі (+) та (-) позначено підвищення або зниження експресії гена між першим та другим триместрами фізіологічної вагітності.

- Гени, пов’язані з обміном стероїдів: (+)HSD11B1, (+)HSD17B2.
- Гени, пов’язані з регулюванням росту та розвитку: (+)BMP5, (-)MT1G, (+)GRB14, (-)GPC6.
- Гени, пов’язані зі структурою клітинної мембрани: (-)EPCAM, (+)TMEM100, (-)TSPAN8.
- Гени, що відіграють різні функції: (-)PAEP, (+)HIGD1B, (+)HECW2-AS1.

Список використаної літератури:

1. Li W. Volcano plots in analyzing differential expressions with mRNA microarrays. J Bioinform Comput Biol. 2012 Dec;10(6):1231003. doi: 10.1142/S0219720012310038
2. Li W, Freudenberg J, Suh YJ, Yang Y. Using volcano plots and regularized-chi statistics in genetic association studies. Comput Biol Chem. 2014 Feb; 48:77-83. doi: 10.1016/j.compbiolchem.2013.02.003
3. Rahmatallah Y, Emmert-Streib F, Glazko G. Gene set analysis approaches for RNA-seq data: performance evaluation and application guideline. Brief Bioinform. 2016 May;17(3):393-407. doi: 10.1093/bib/bbv069
4. Lykhenko O. K., Frolova A. O., Obolenskaya M. Yu. Changes in the human placental transcriptome during the physiological course of pregnancy. Biopolym. Cell. 2021 Feb;37(1):73-82. doi:10.7124/bc.000A4D
5. GitHub - hbctraining/Training-modules: Materials for short, half-day workshops. GitHub. URL: <https://github.com/hbctraining/Training-modules>
6. Xiao Y, Hsiao TH, Suresh U, et al. A novel significance score for gene selection and ranking. Bioinformatics. 2014;30(6):801-807. doi:10.1093/bioinformatics/btr671