

Функціональна гетерогенність фібробластів у шкірі людини: одноклітинний транскриптомний аналіз та вікові зміни

Team 3: Oleksandr Surma, Anastasiia Bliuma, Nikita Oleinikov

Стаття та датасет, який досліджували

communications biology

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [communications biology](#) > [articles](#) > [article](#)

Article | [Open access](#) | Published: 23 April 2020

Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming

Llorenç Solé-Boldo, Günter Raddatz, Sabrina Schütz, Jan-Philipp Malim, Karsten Rippe, Anke S. Lonsdorf, Manuel Rodríguez-Paredes & Frank Lyko

Communications Biology 3, Article number: 188 (2020) | [Cite this article](#)

59k Accesses | 22 Altmetric | [Metrics](#)

Набори даних scRNA-seq доступні в базі даних Gene Expression Omnibus (GEO)

(номер доступу [GSE130973](#))

Supplementary file	Size	Download	File type/resource
GSE130973_barcodes_filtered.tsv.gz	66.9 Kb	(ftp)(http)	TSV
GSE130973_barcodes_raw.tsv.gz	10.9 Mb	(ftp)(http)	TSV
GSE130973_genes_filtered.tsv.gz	251.2 Kb	(ftp)(http)	TSV
GSE130973_genes_raw.tsv.gz	251.2 Kb	(ftp)(http)	TSV
GSE130973_matrix_filtered.mtx.gz	62.0 Mb	(ftp)(http)	MTX
GSE130973_matrix_raw.mtx.gz	137.9 Mb	(ftp)(http)	MTX
GSE130973_seurat_analysis_lyko.rds.gz	441.0 Mb	(ftp)(http)	RDS

Об'єкт дослідження

~20,000 клітин були отримані зі здорових зразків цілої шкіри відразу після резекції пахово-клубової області



Після застосування фільтрів призвело до остаточного набору даних із 15 457 одноклітинних транскриптомів.

 OSMOSIS
from ELSEVIER

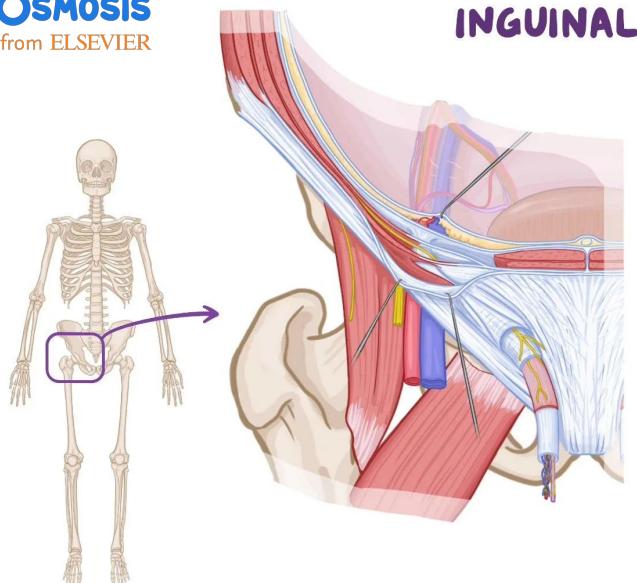


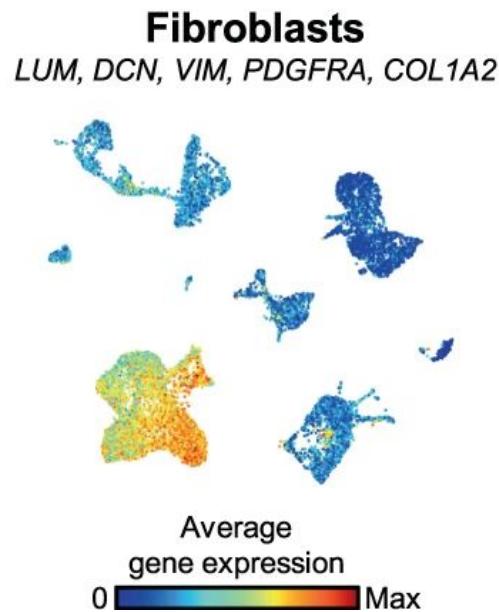
Table 1. Overview of the samples used for single-cell RNA sequencing.

The table shows biological features and sequencing statistics for the five samples.

Sample ID	Gender	Age	Skin type	Reads per sample	Numbers of cells	Reads per cell	Genes per cell
Donor 1	Male	25	Fair	322,091,192	2784	102,904	1343
Donor 2	Male	27	Fair	338,738,780	2670	119,737	1111
Donor 3	Male	53	Fair	359,776,321	3324	107,976	1718
Donor 4	Male	70	Fair	378,219,220	2144	170,215	1388
Donor 5	Male	69	Fair	370,342,531	4535	81,411	872

Основні знахідки.

Було ідентифіковано чотири основні субпопуляції фібробластів, які мають просторову локалізацію та відрізняються функціональністю:



Secretory_papillary (в нашому аналізі - Fibroblasts / Pericytes)

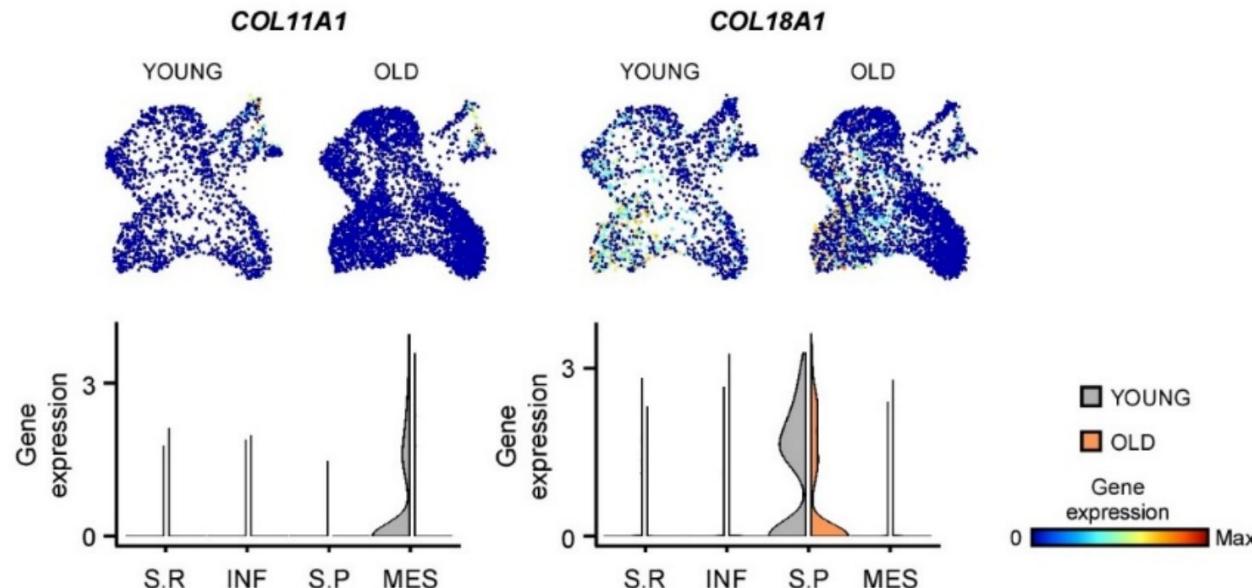
Secretory_reticular (в нашому аналізі - Matrix fibroblasts)

Pro_inflammatory (в нашому аналізі - Immune-regulatory fibroblasts)

Mesenchymal (в нашому аналізі - Neuroendocrine / Mesenchymal)

Основні знахідки.

Після - порівнювали експресію генів між молодими та старими фібробластами в кожному кластері. Наприклад:



Важливі висновки статті.

1. З віком спостерігається зниження “праймінгу” фібробластів — тобто їхньої функціональної спеціалізації.
2. Кожен підтип фібробласта мав унікальний профіль експресії генів у молодості. У літніх людей відмінності між цими підтипами розмиті - їх профілі стають схожими.
3. Було виявлено значне зменшення передбачених взаємодій між дермальними фібробластами та іншими клітинами шкіри, включаючи недиференційовані кератиноцити на межі дерми та епідермісу.

Ці результати свідчать про втрату клітинної ідентичності як важливу вікову зміну в дермі людини.

Завдання нашого аналізу. Блок 1.

1. Створити SeuratObject з raw data.
2. Провести покроковий QC, але більш детальний, ніж в статті, провівши додаткові кроки.
3. Виконати кластеризацію та побудову UMAP + зробити анотацію всіх кластерів.
4. Ідентифікувати субтипи фібробластів та їх біологічні функції (GO).

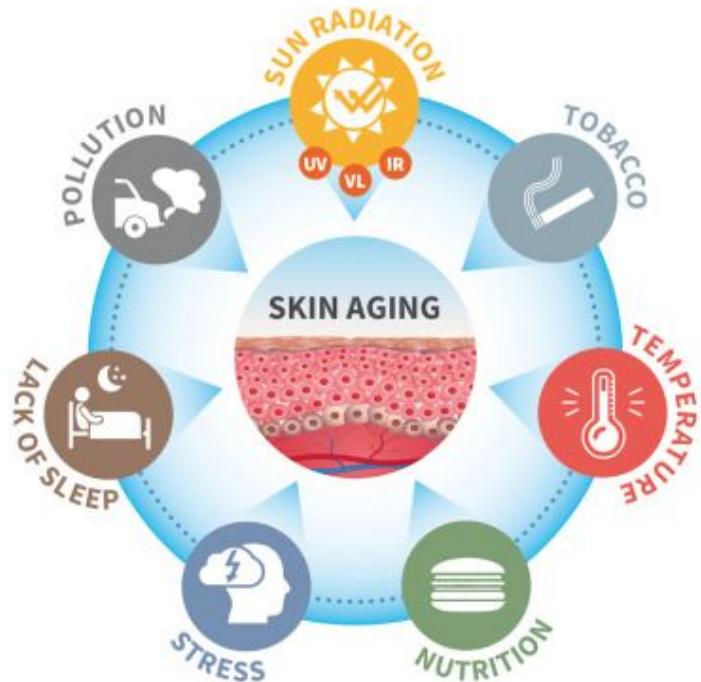
Завдання нашого аналізу. Блок 2.

1. Повторити основні кроки препроцесингу на SeuratObject зі статті.
2. Аnotувати фібробластні субтипи за маркерними генами.
3. Порівняти експресію генів між молодими та літніми донорами для кожного підтипу фібробластів (Young vs Old).

Дизайн експерименту.

Дослідники зібрали зразки шкіри з ділянки, захищеної від сонця, у здорових донорів різного віку.

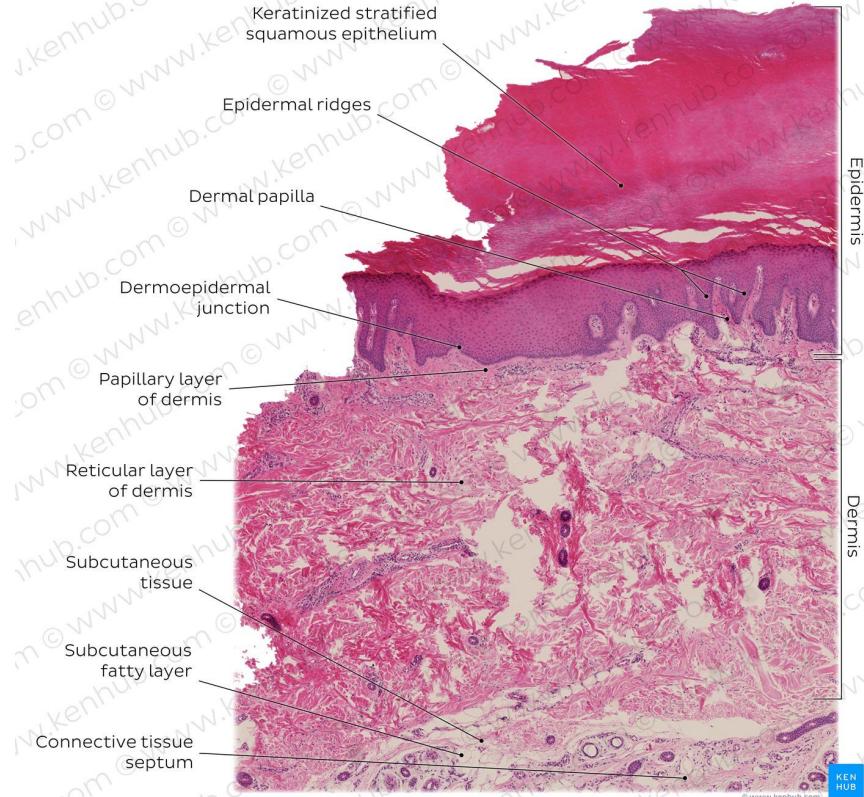
Це дозволило мінімізувати вплив зовнішніх факторів, таких як ультрафіолетове випромінювання, на результати дослідження.



DOI: [10.1016/j.jdermsci.2016.09.015](https://doi.org/10.1016/j.jdermsci.2016.09.015)

Підготовка зразків

1. Спочатку брали невеликий шматочок шкіри (~ 3-4 мм).
2. Шкіру розділили на епідерміс і дерму.



Підготовка зразків

3. З дерми виділили живі клітини шляхом ферментативного розщеплення.

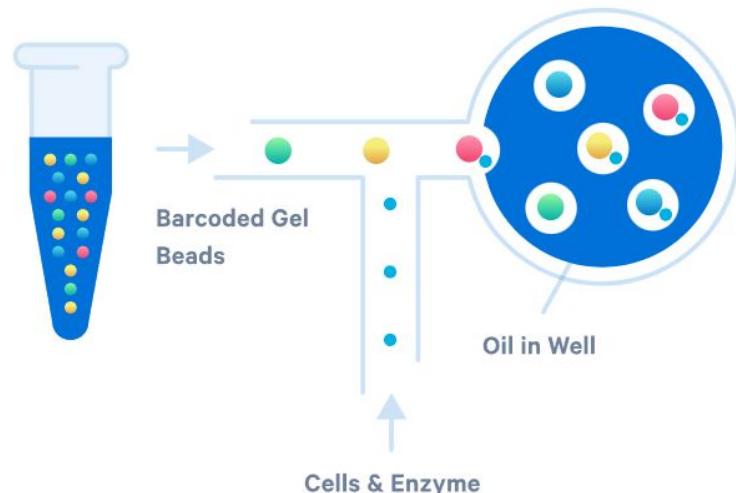


Whole Skin Dissociation Kit, human

The Whole Skin Dissociation Kit, human enables the gentle and efficient generation of single-cell suspensions from human skin tissue without separating dermis and epidermis. The kit has been particularly developed for the isolation of fibroblasts from diverse human skin biopsies as well as for phenotyping or enumeration of human skin cell populations by flow cytometry.

Підготовка зразків

4. Використали 10x Genomics **Chromium Single Cell Controller** - для захоплення і маркування окремих клітин.

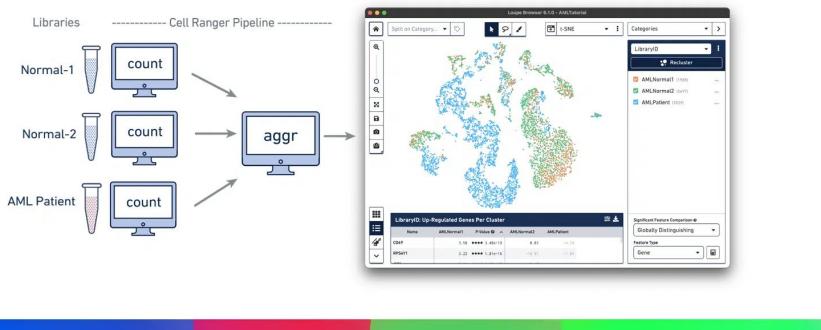


5. Секвенування було виконано за допомогою пристрою **HiSeq 4000 (Illumina)**.

illumina®

Підготовка зразків

10XGenomics/
cellranger



Бібліотеки, які використовувались в проекті.

library(tibble) - Робота з таблицями

library(gt) - Створення гарних таблиць

library(Seurat) - Основний пакет для аналізу
scRNA-seq даних

library(tidyverse) - Для роботи з даними

library(SingleCellExperiment) - Альтернативна
структуря даних для scRNA-seq

library(scDblFinder) - Для виявлення doublets

library(scater) - QC, нормалізація та візуалізація
даних у SingleCellExperiment

library(clustree) - Візуалізація кластерних дерев при
різних параметрах кластеризації

library(cluster) - Для обчислення silhouette scores

library(clusterProfiler) - GO аналіз

library(org.Hs.eg.db) - Анотаційна БД для людини.

library(presto) - Для швидкого диференційного
аналізу

library(plotly) - Інтерактивна графіка

ПЛАН ПРОЄКТУ

Блок 1. Повна обробка та створення SeuratObject з нуля.

Блок 2. Аналіз вже створеного SeuratObject зі статті.

Висновки.

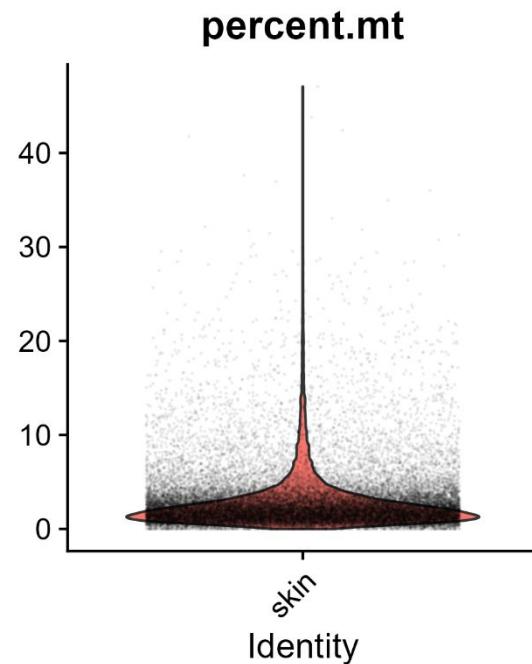
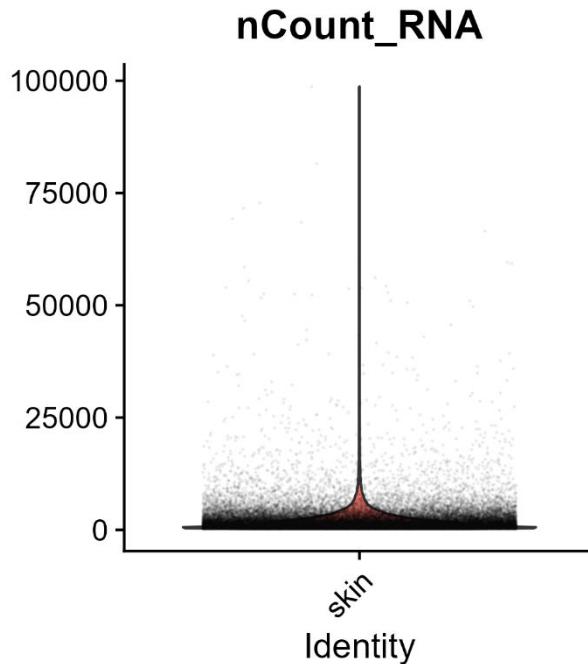
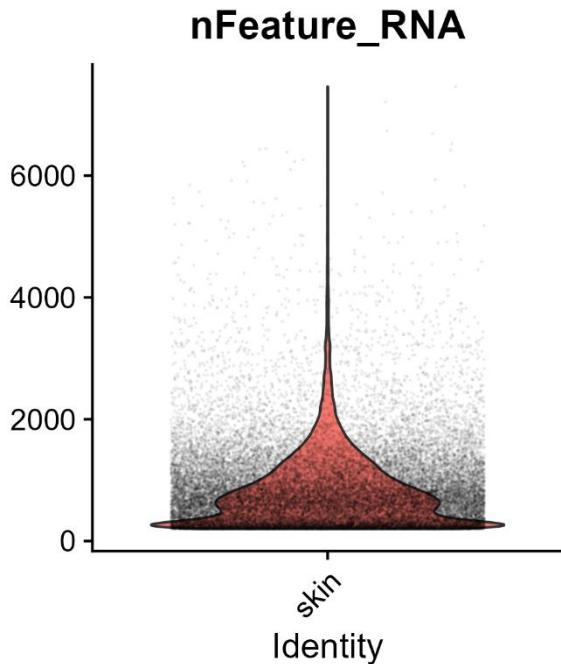
Блок 1. Повна обробка та створення SeuratObject з нуля

Data: використовуючи Seurat::Read10X(), було створено SeuratObject, який вже мав інтегровані датасети (2 Young + 3 Old).

Supplementary file	Size	Download	File type/resource
GSE130973_barcodes_filtered.tsv.gz	66.9 Kb	(ftp)(http)	TSV
GSE130973_barcodes_raw.tsv.gz	10.9 Mb	(ftp)(http)	TSV
GSE130973_genes_filtered.tsv.gz	251.2 Kb	(ftp)(http)	TSV
GSE130973_genes_raw.tsv.gz	251.2 Kb	(ftp)(http)	TSV
GSE130973_matrix_filtered.mtx.gz	62.0 Mb	(ftp)(http)	MTX
GSE130973_matrix_raw.mtx.gz	137.9 Mb	(ftp)(http)	MTX
GSE130973_seurat_analysis_lyko.rds.gz	441.0 Mb	(ftp)(http)	RDS

Блок 1. Pre Quality Control

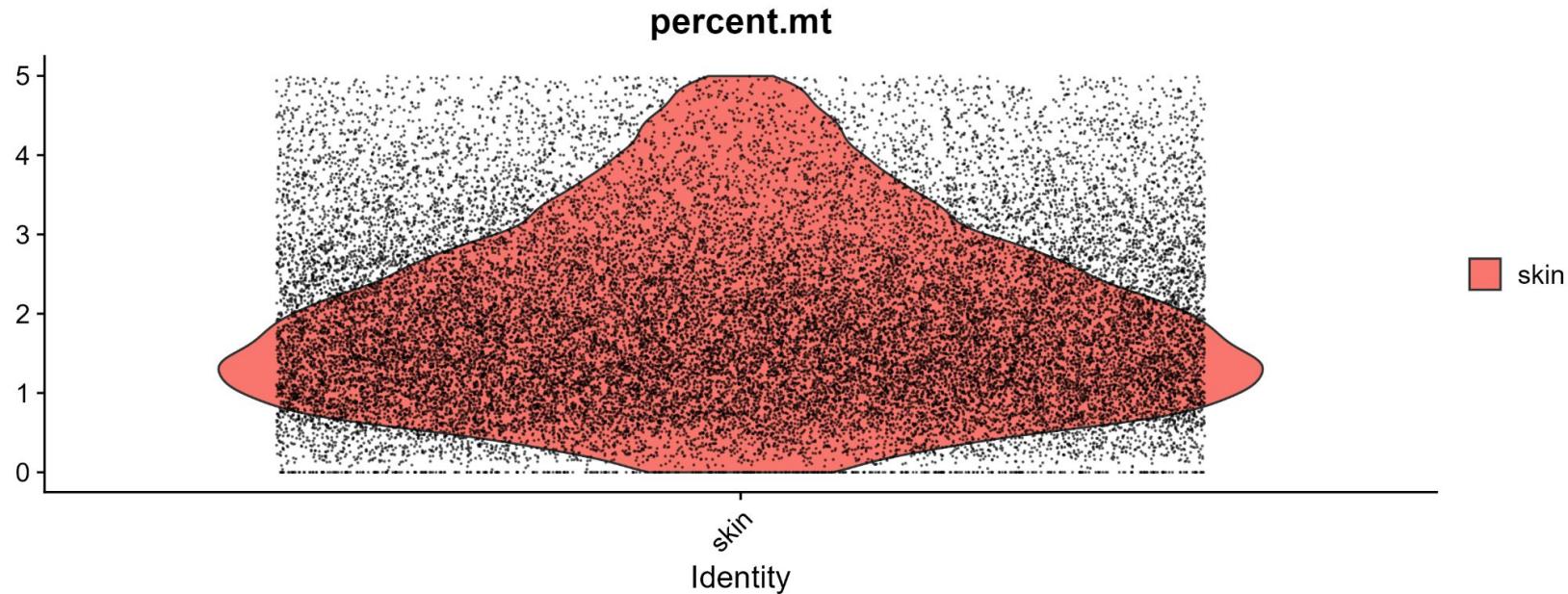
Подивимось на VlnPlot до quality control.



Блок 1. Quality Control: mt5 or mt10?

Подивимось як впливає вибір відсотку мітохондріальних генів, що буде видалено, на якість та кількість клітин.

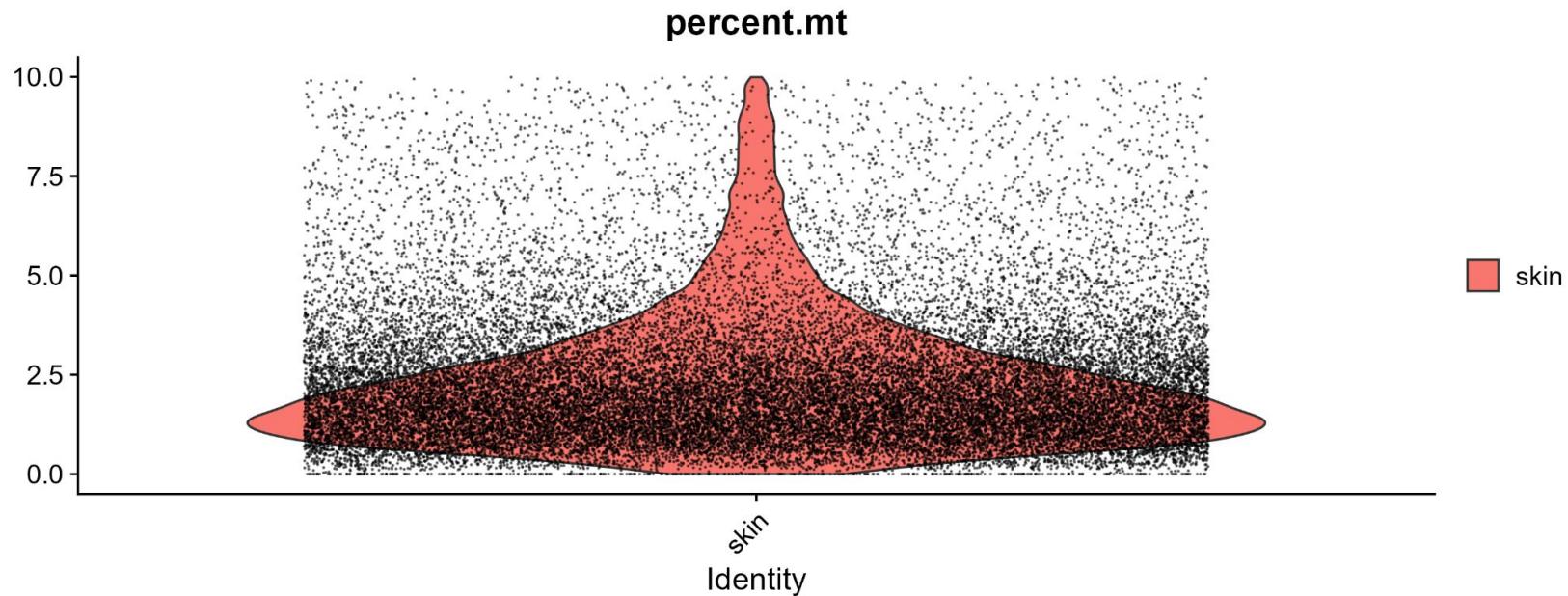
Кількість клітин: 29334



Блок 1. Quality Control: mt5 or mt10?

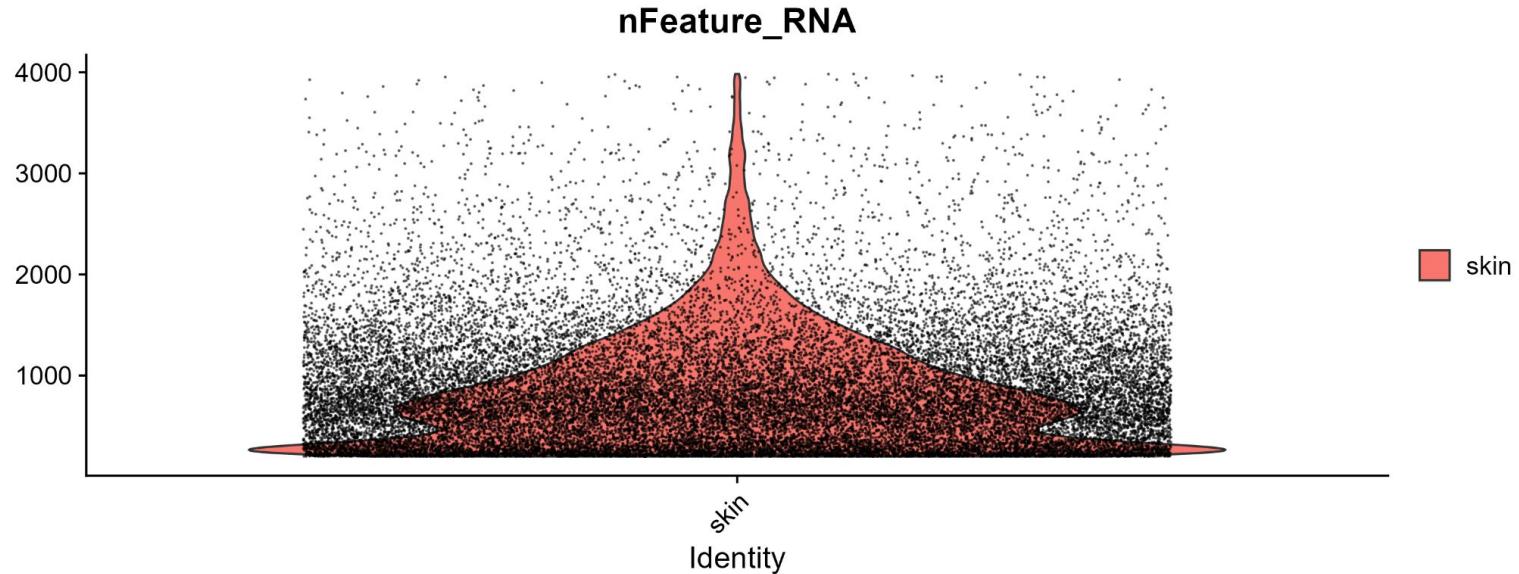
Подивимось як впливає вибір відсотку мітохондріальних генів, що буде видалено, на якість та кількість клітин.

Кількість клітин: 32033



Блок 1. Quality Control: nFeature_RNA.

Проводимо QC, видаляючи всі об'єкти, що містять більше 4000 експресованих генів.



Блок 1. SCTransform() + RunPCA()

Для подальшого QC і загалом роботи з датасетом проводимо нормалізацію та масштабування за допомогою загальної лінійної моделі (GLM). Після, запускаємо аналіз головних компонент (PCA).

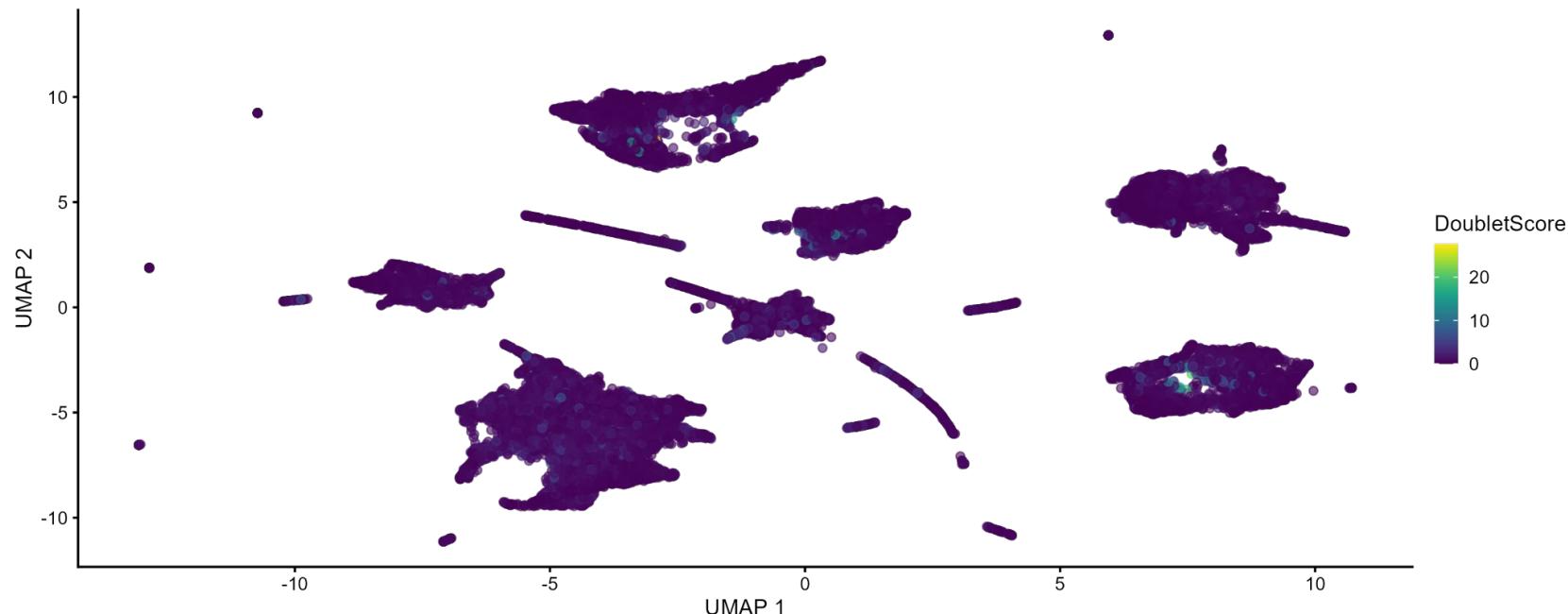
```
seurat_obj <- SCTransform(seurat_obj, verbose = FALSE)
```

```
seurat_obj <- RunPCA(seurat_obj, features = VariableFeatures(object =  
seurat_obj))
```

Блок 1. Quality Control: scDblFinder

Проводимо QC видаляючи даблети за допомогою пакету scDblFinder.

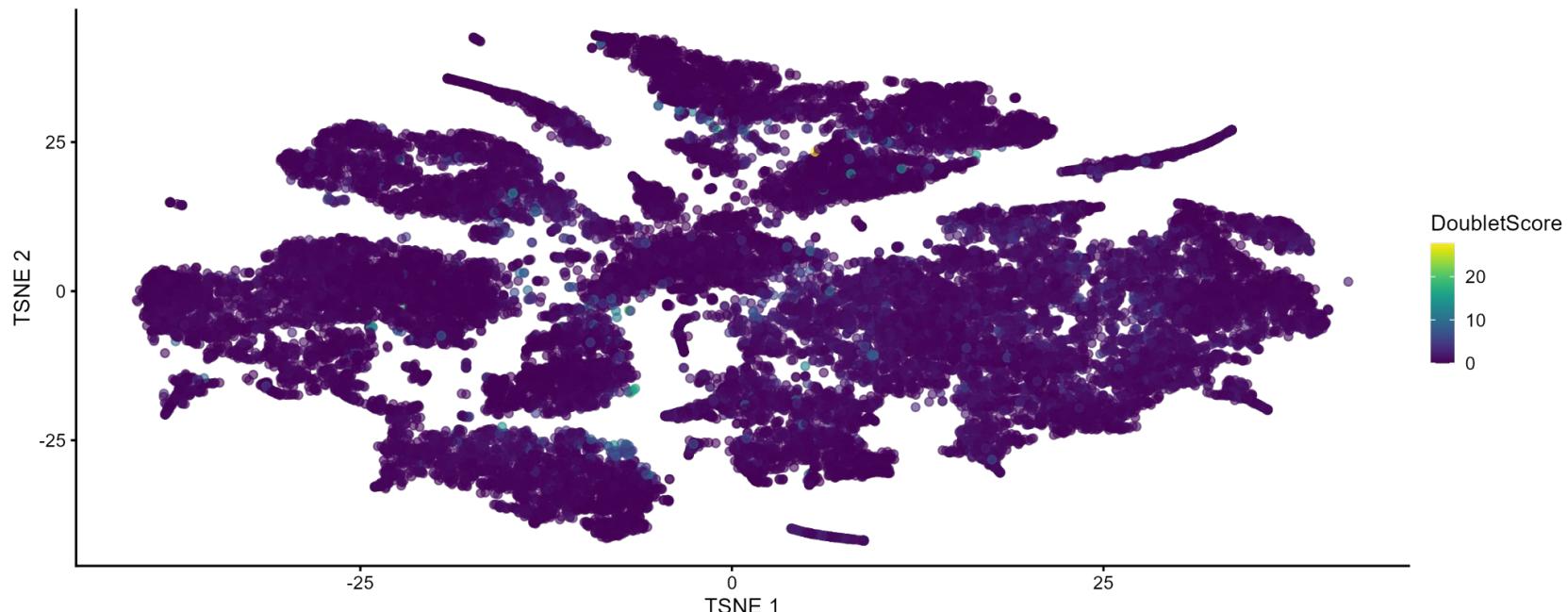
Створюємо UMAP-візуалізацію на основі PCA.



Блок 1. Quality Control: scDbIFinder

Проводимо QC видаляючи даблети за допомогою пакету scDbIFinder.

В додаток створимо TSNE-візуалізацію на основі PCA

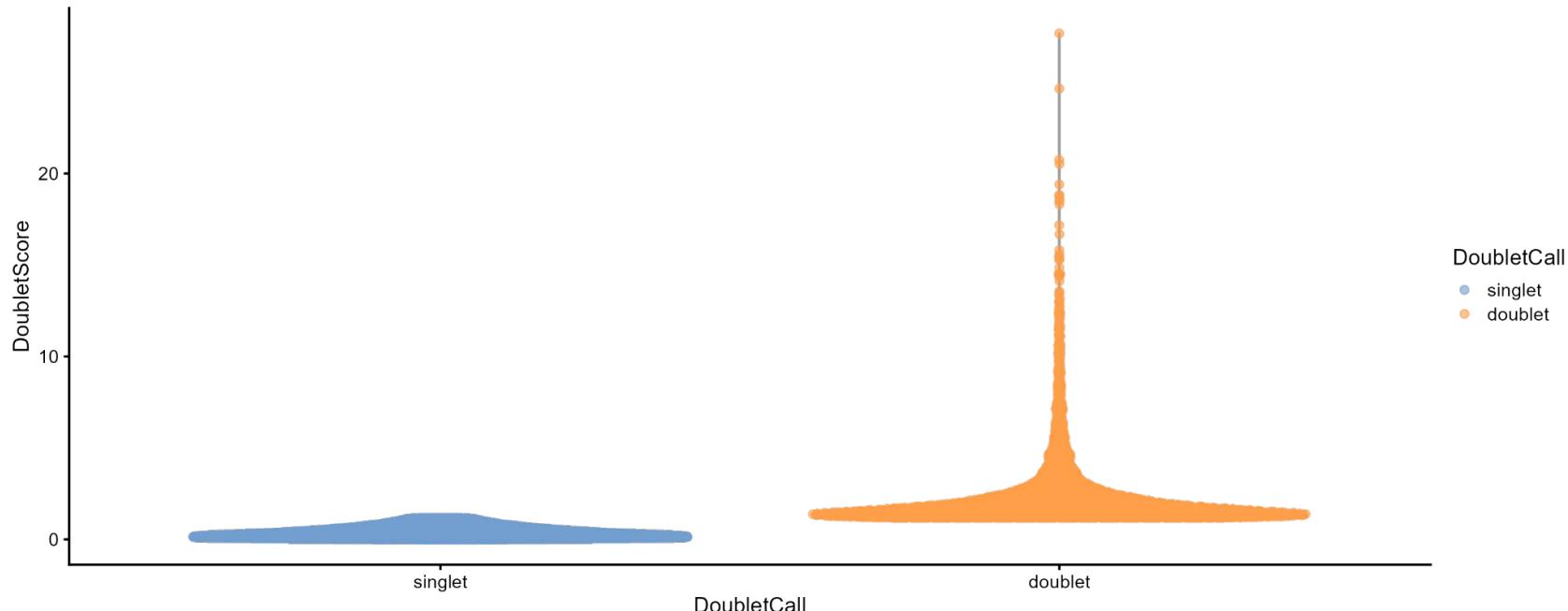


Блок 1. Quality Control: scDbIFinder

Подивимось, скільки singlet, скільки doublet

singlet	doublet
24881	4306

Подивимось на `plotColData`, який показує розподіл `DoubletScore` між групами 'singlet' і 'doublet'



Блок 1. Quality Control

1. Первинний QC - створюючи SeuratObject ми вже прибрали клітини, які експресують менше 200 генів, а також залишили тільки ті гени, які присутні хоча б у 3 клітинах.
2. Видалили всі об'єкти, що містять більше 5% мітохондріальних генів та більше 4000 експресованих генів.
3. Видалили даблети за допомогою пакету scDblFinder (стратегія полягає в симуляцію *in silico* дублетів із одноклітинних профілів експресії).

РЕЗУЛЬТАТ: БУЛО:

```
dim(seurat_obj_raw)  
21975 33175
```

СТАЛО:

```
dim(seurat_clean)  
20332 24881
```

Блок 1. Pre_clustering

Проводимо первинну кластеризацію, продивляючись можливі варіанти кластерів.

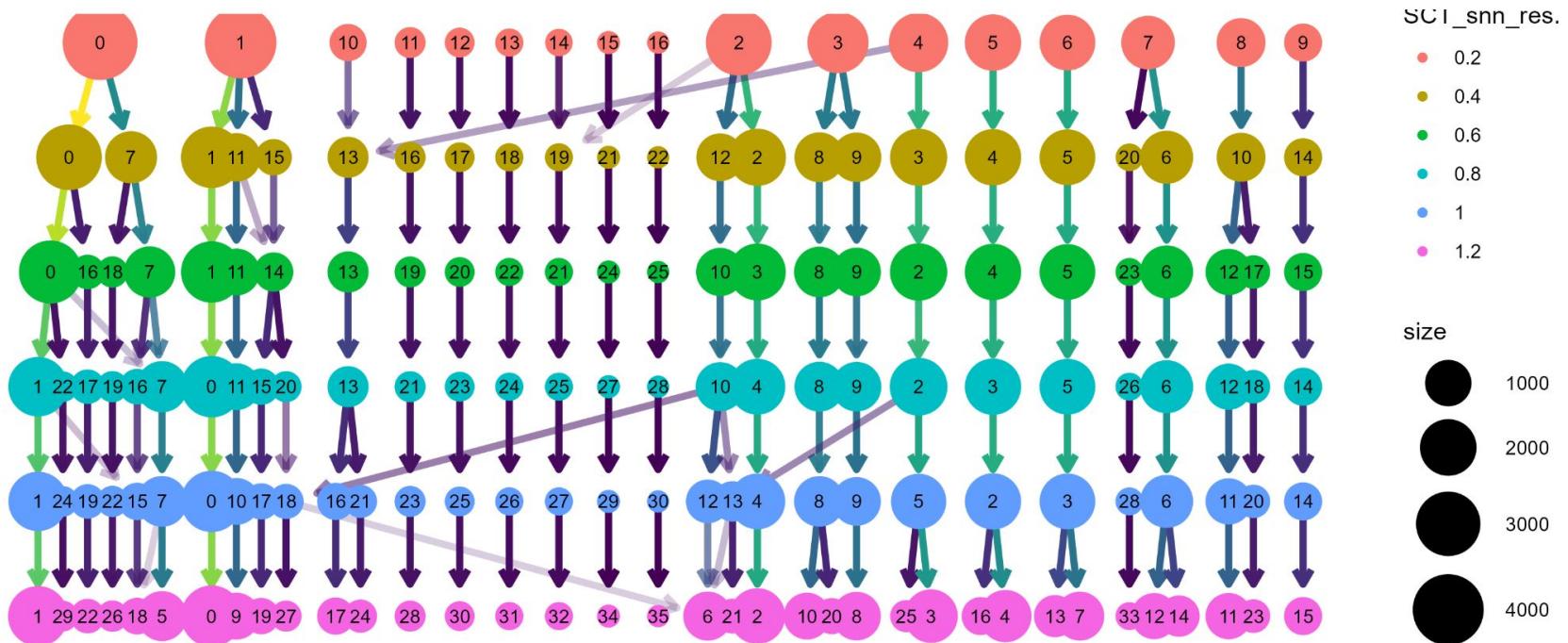
```
seurat_clean <- FindNeighbors(seurat_clean, dims = 1:30)  
resolutions <- seq(from = 0.2, to = 1.2, by = 0.2)  
FindClusters(seurat_clean, resolution = resolutions)
```



```
RunUMAP(seurat_clean, dims = 1:30)
```

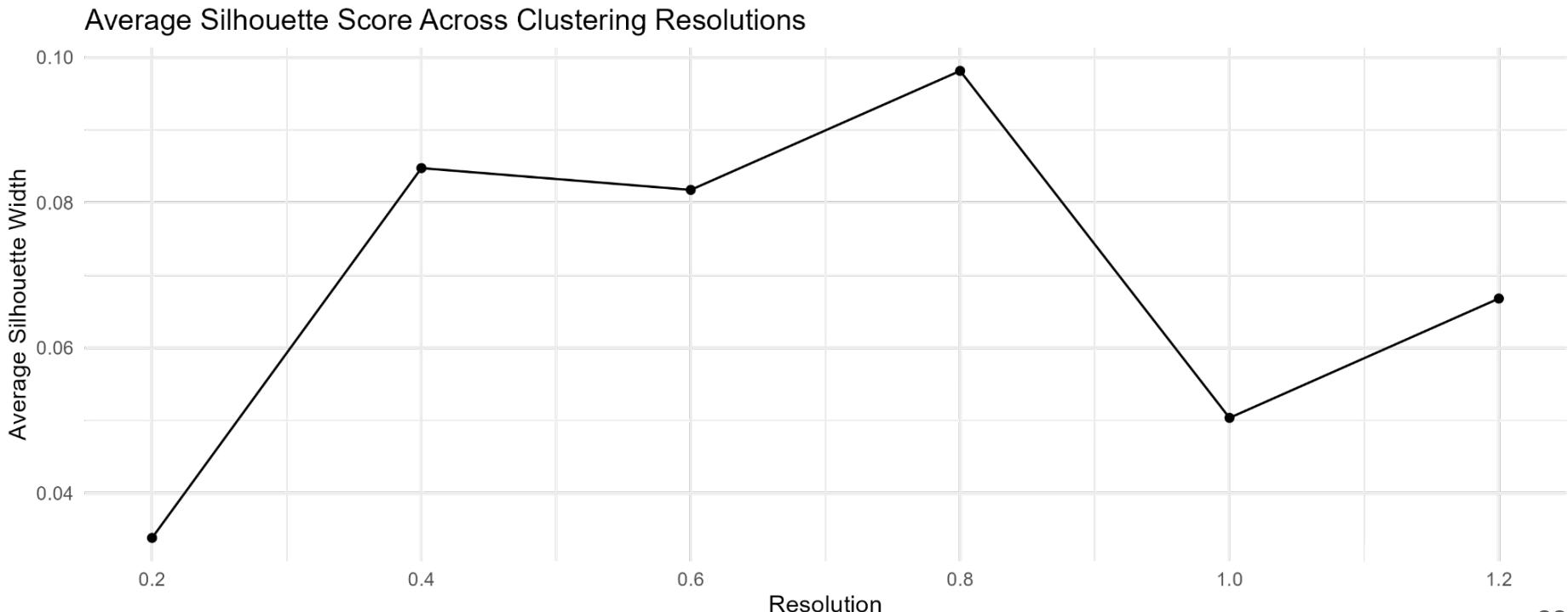
Блок 1. Pre_clustering: clustree

Намалюємо дерево кластерів за допомогою clustree, щоб оцінити стабільність у різних роздільних здатностях.



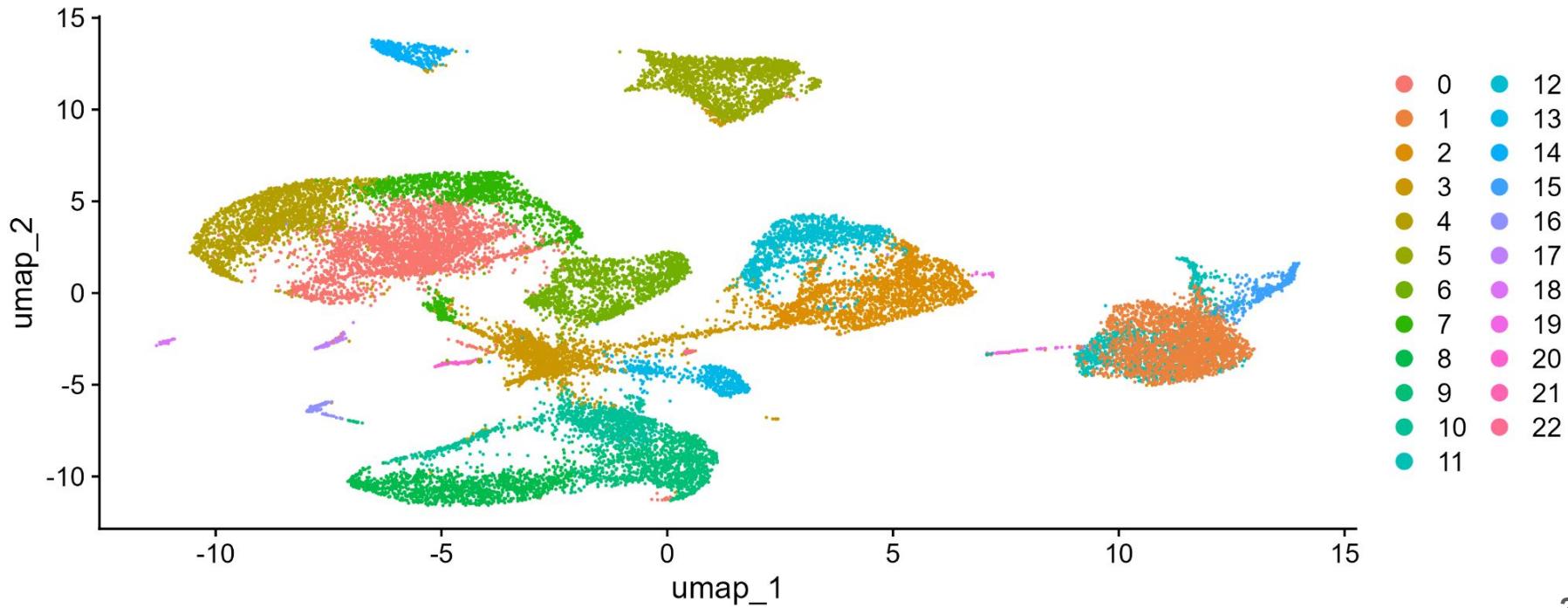
Блок 1. Pre_clustering: silhouette scores

Обчислюємо silhouette scores для кожної роздільної здатності кластеризації



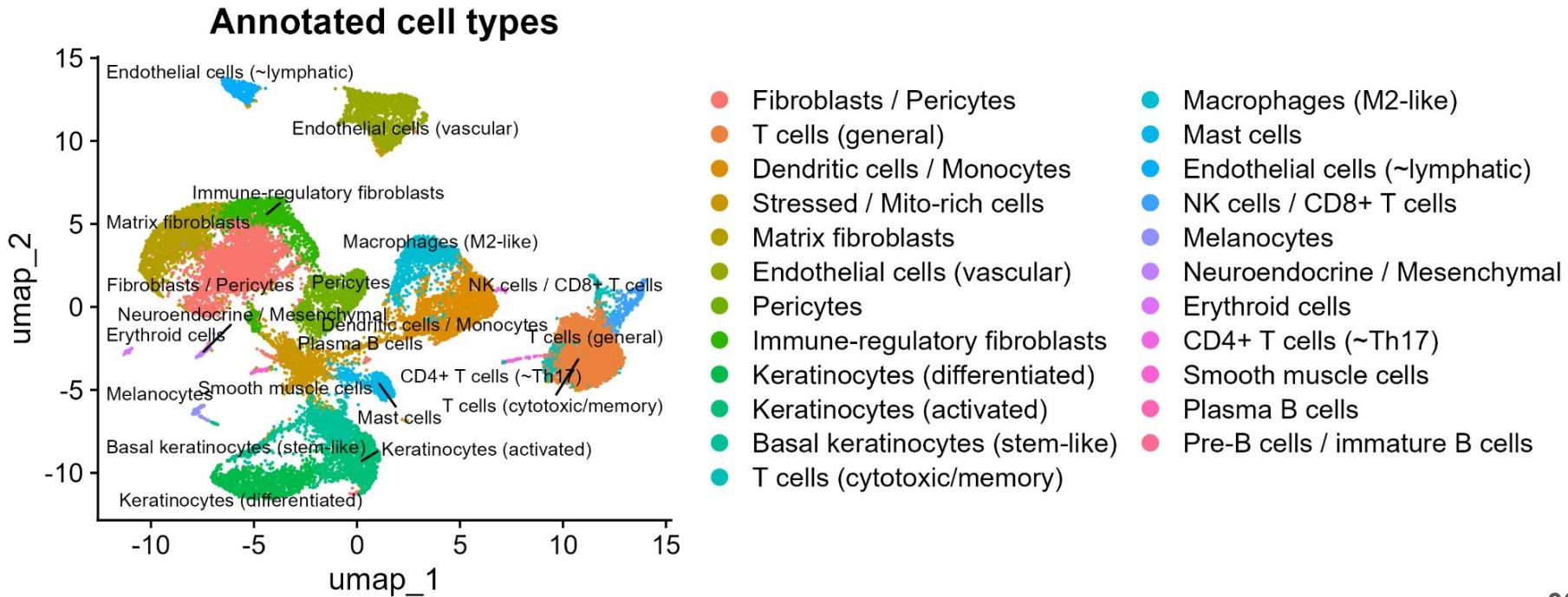
Блок 1. Clustering: UMAP

RunUMAP з resolution = 0.4 та dims = 1:30



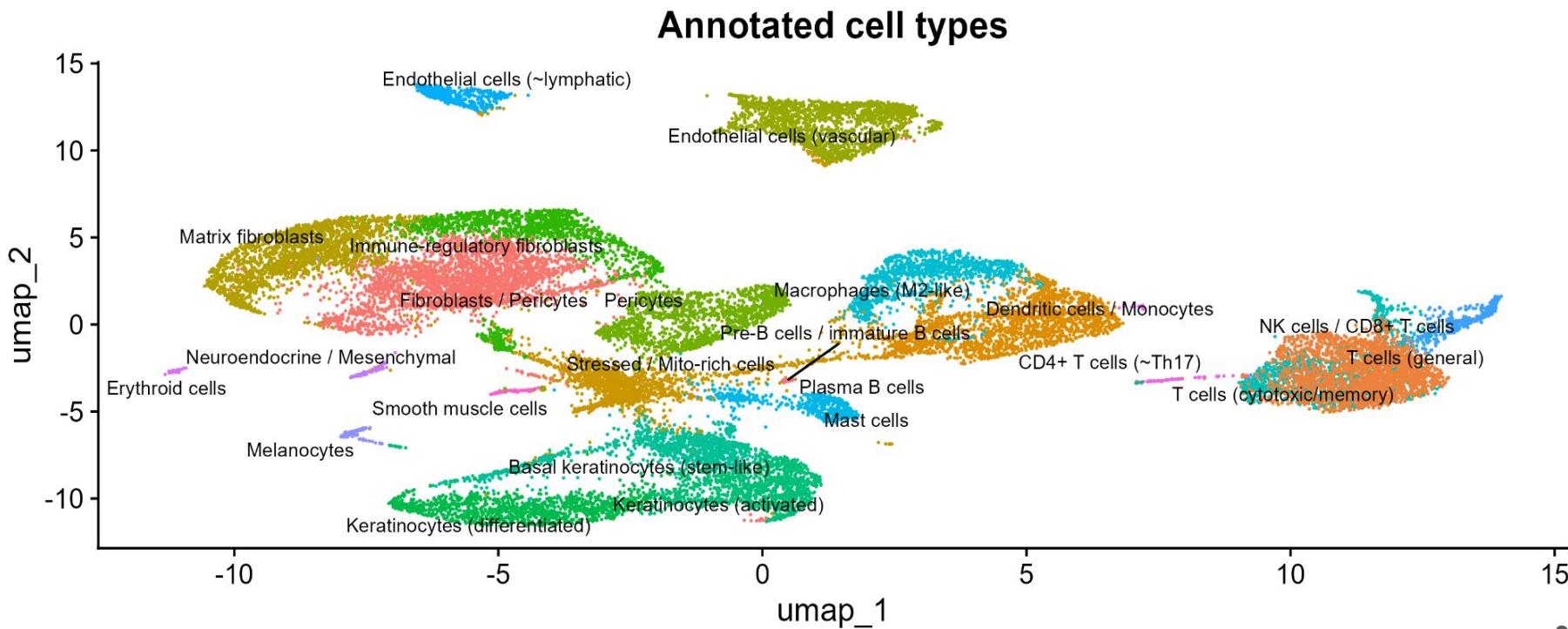
Блок 1. Clustering: annotation

Анотуємо кластери на основі маркерних генів (топ 5)



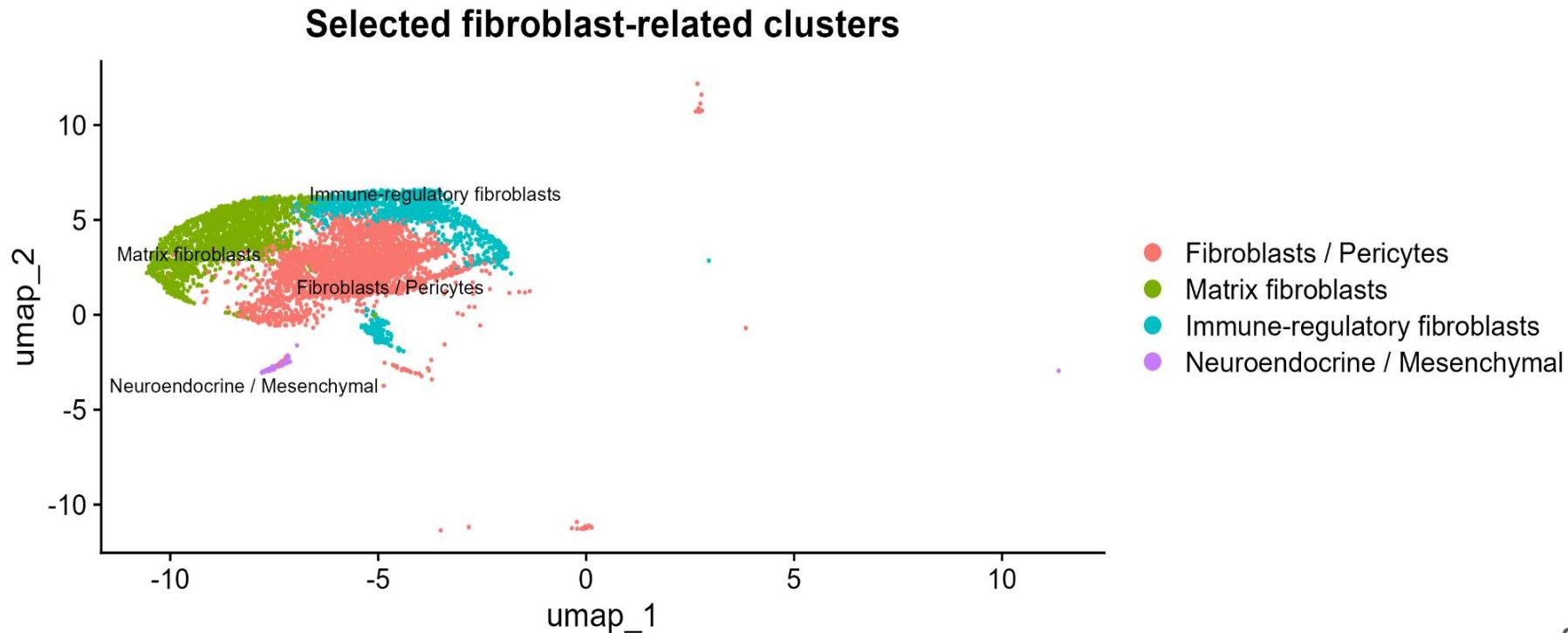
Блок 1. Clustering: annotation

Анотуємо кластери на основі маркерних генів (топ 5)

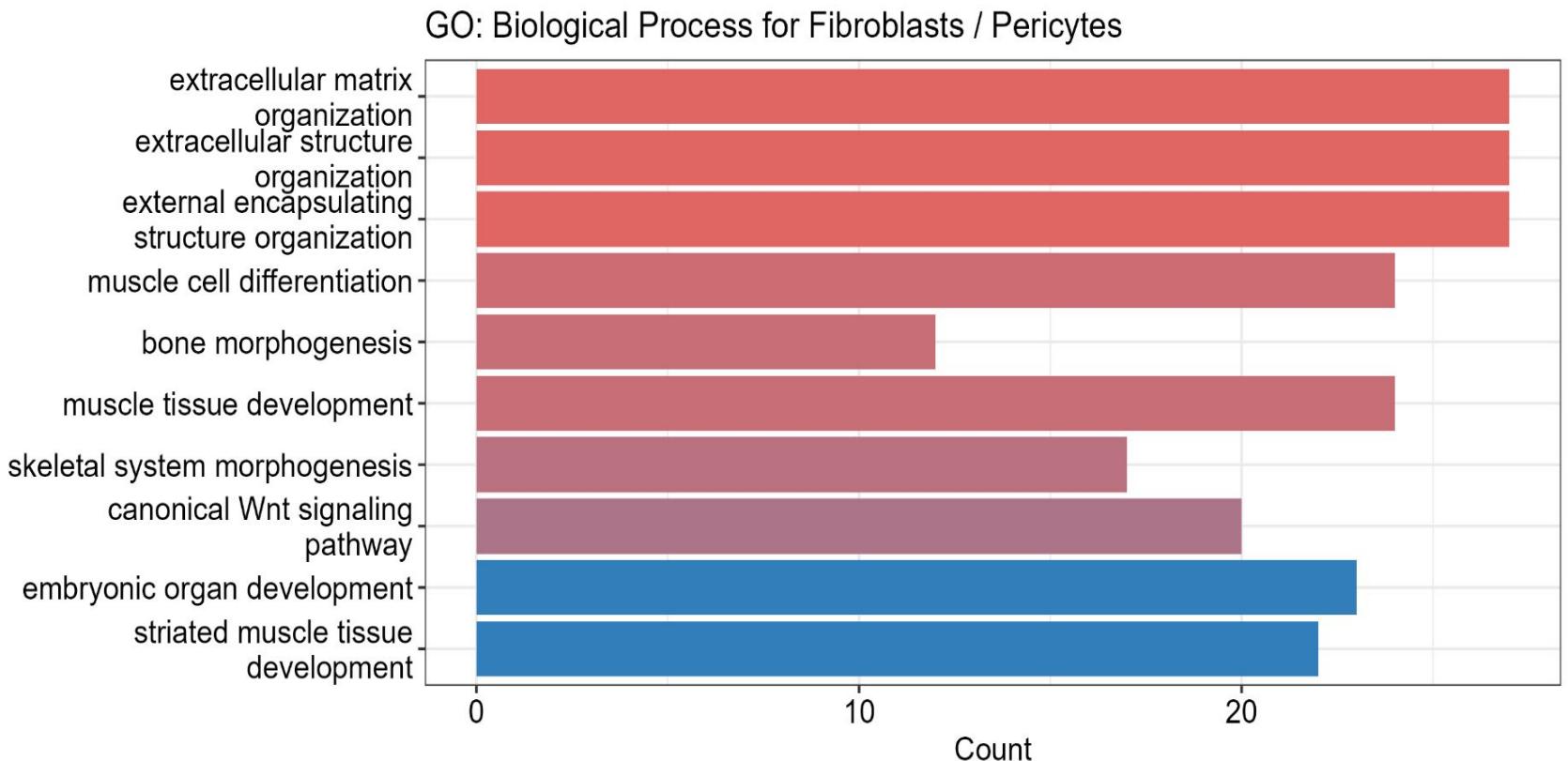


Блок 1. Clustering: fibroblasts

Відокремлюємо fibroblasts окремо від інших типів клітин.

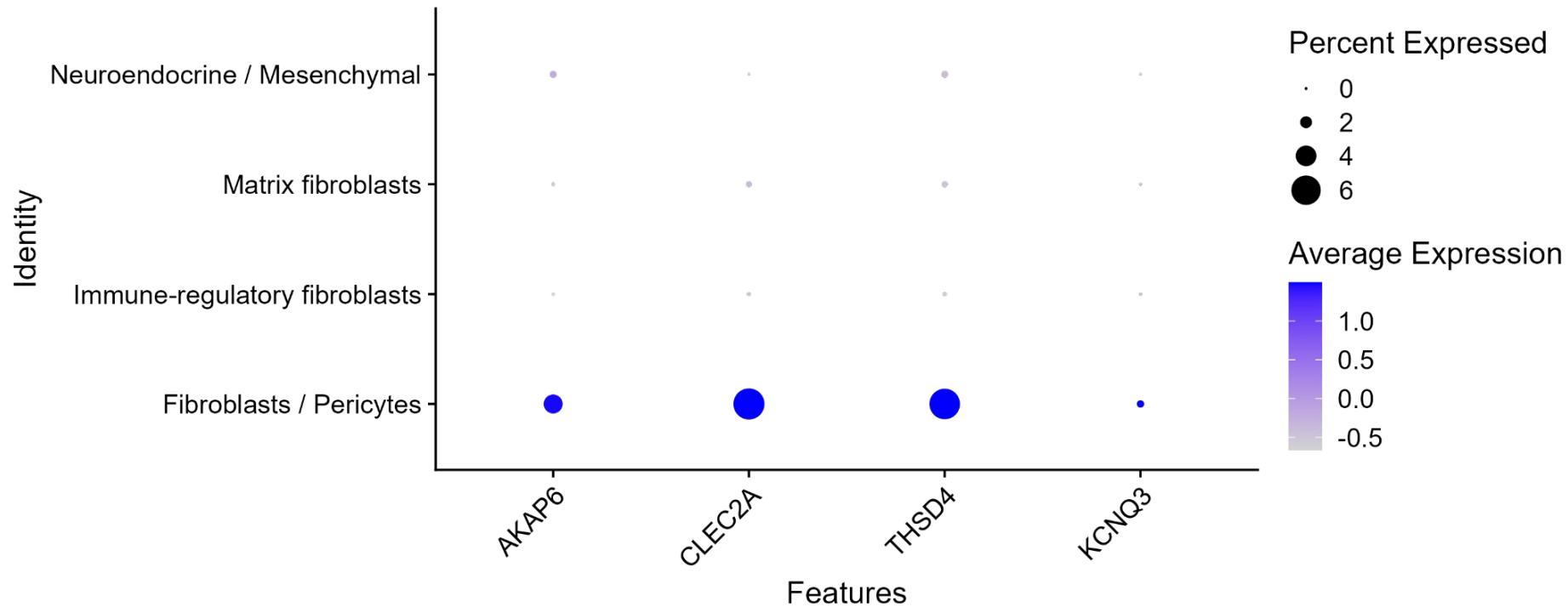


Блок 1.GO: Biological Process for: Fibroblasts / Pericytes.

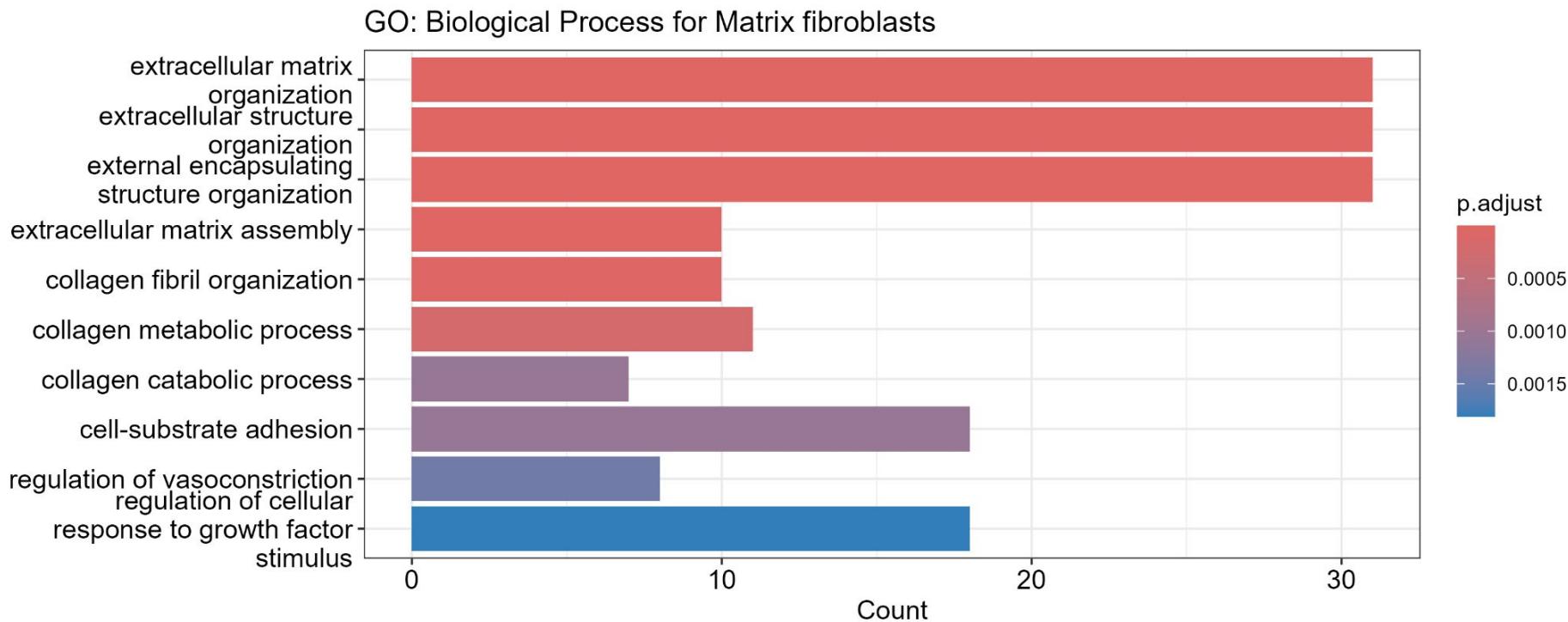


Топ-маркерні гени визначають гетерогенність фібробластів.

Fibroblasts / Pericytes

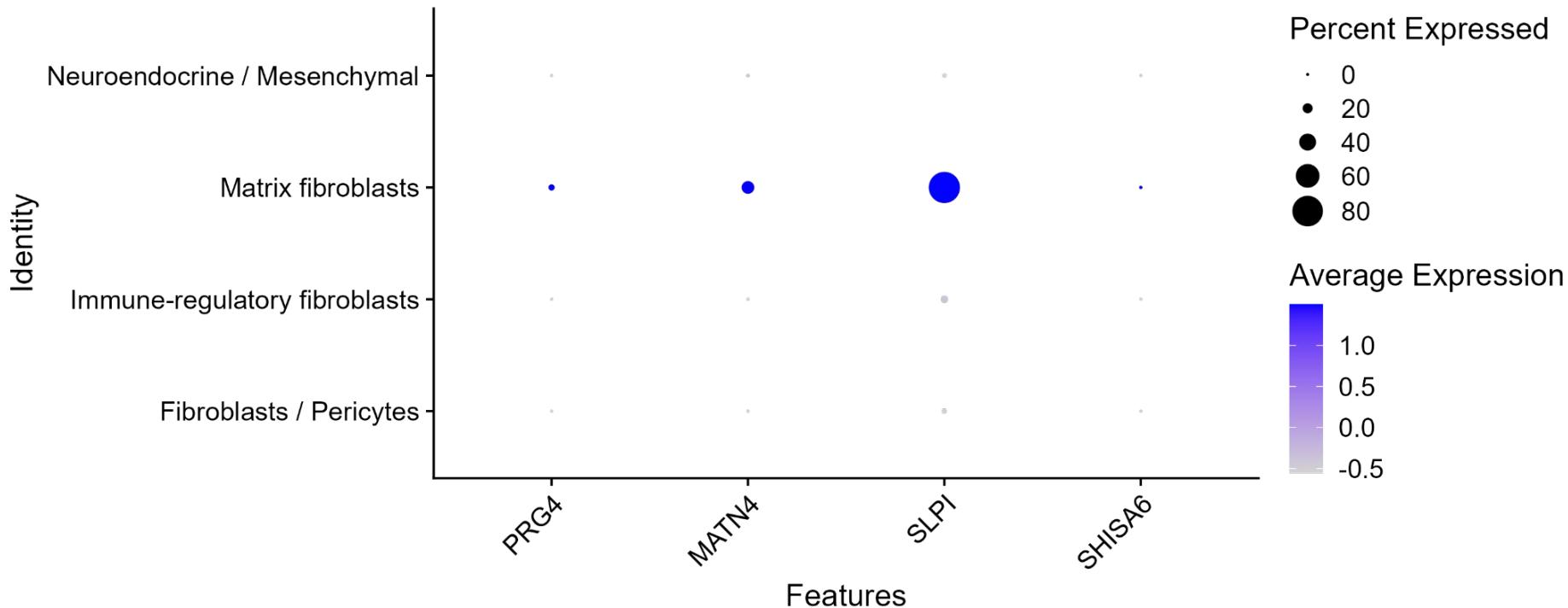


Блок 1.GO: Biological Process for: Matrix fibroblasts.

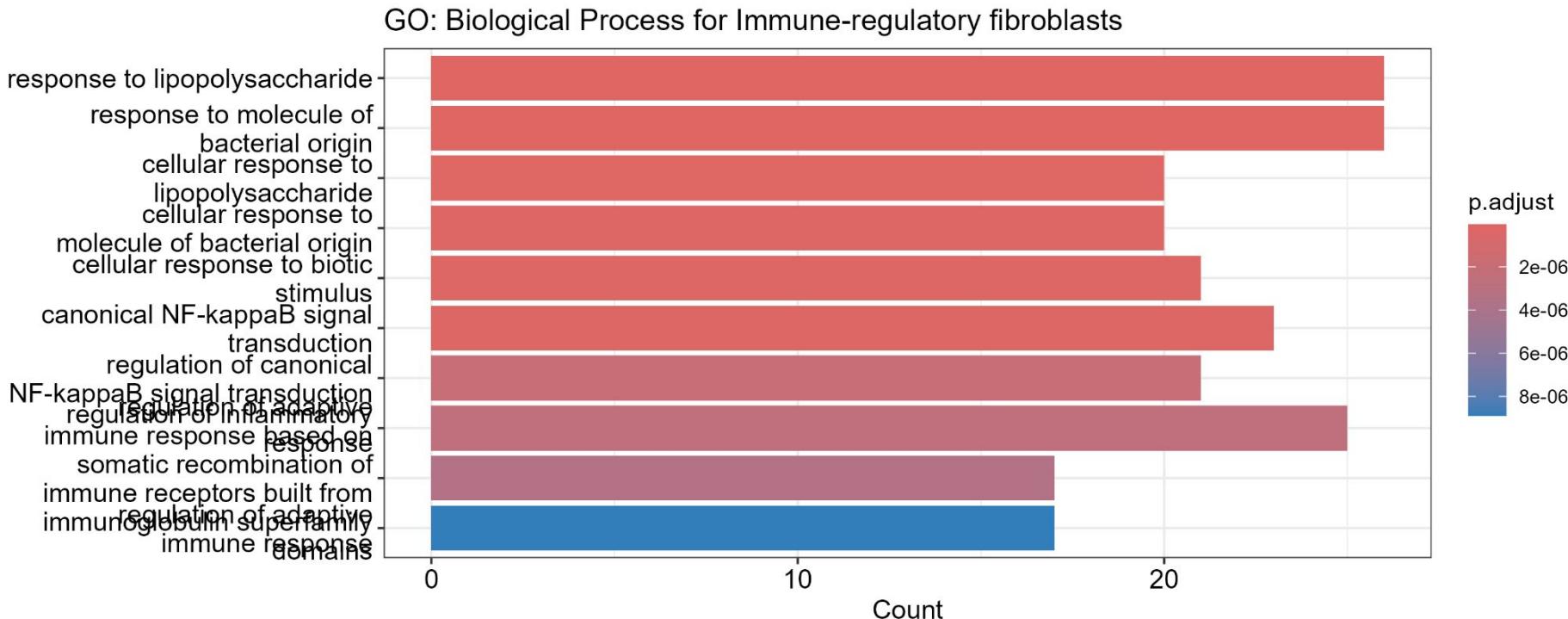


Топ-маркерні гени визначають гетерогенність фібробластів.

Matrix fibroblasts

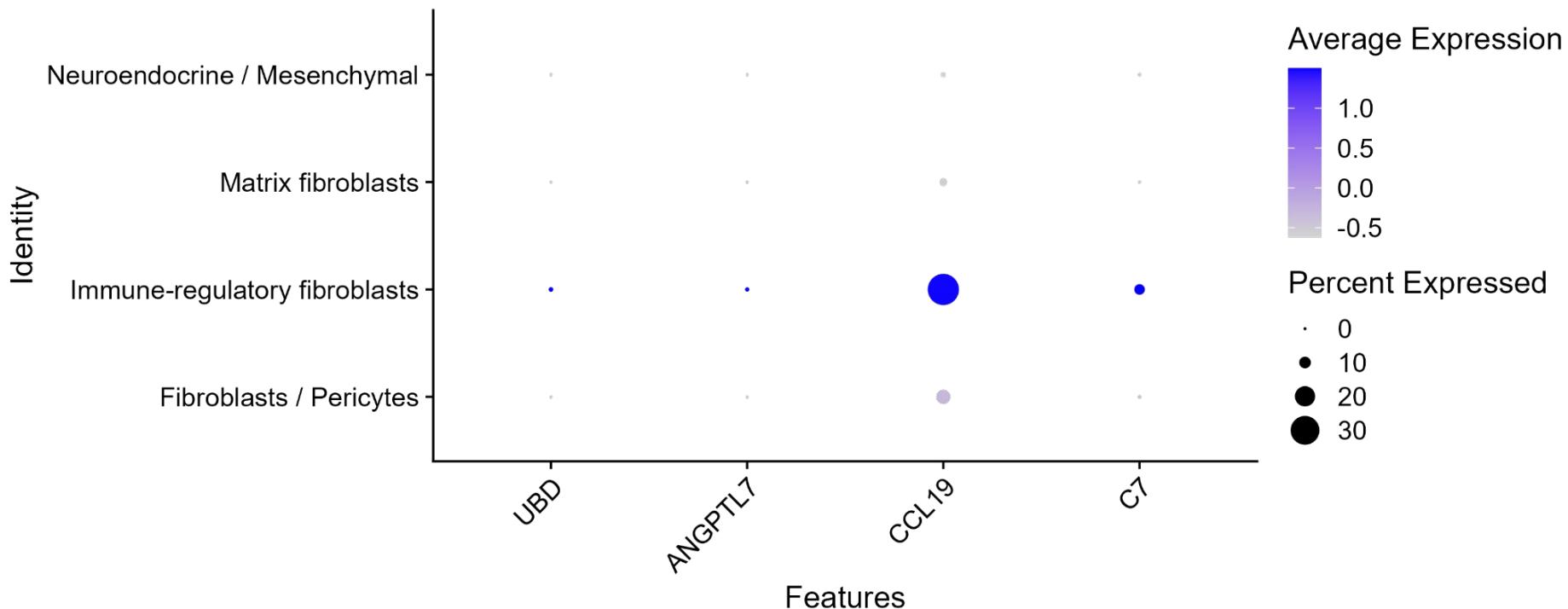


Блок 1.GO: Biological Process for: Immune-regulatory fibroblasts.

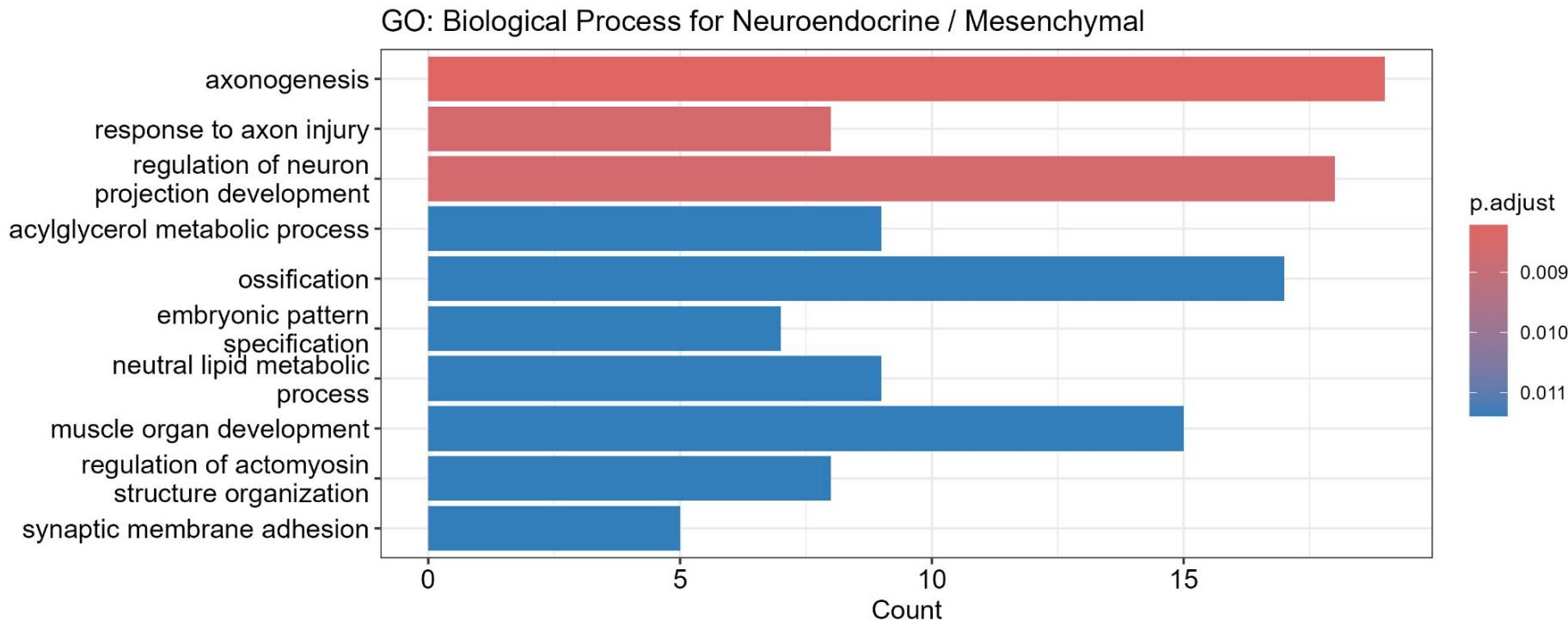


Топ-маркерні гени визначають гетерогенність фібробластів.

Immune-regulatory fibroblasts

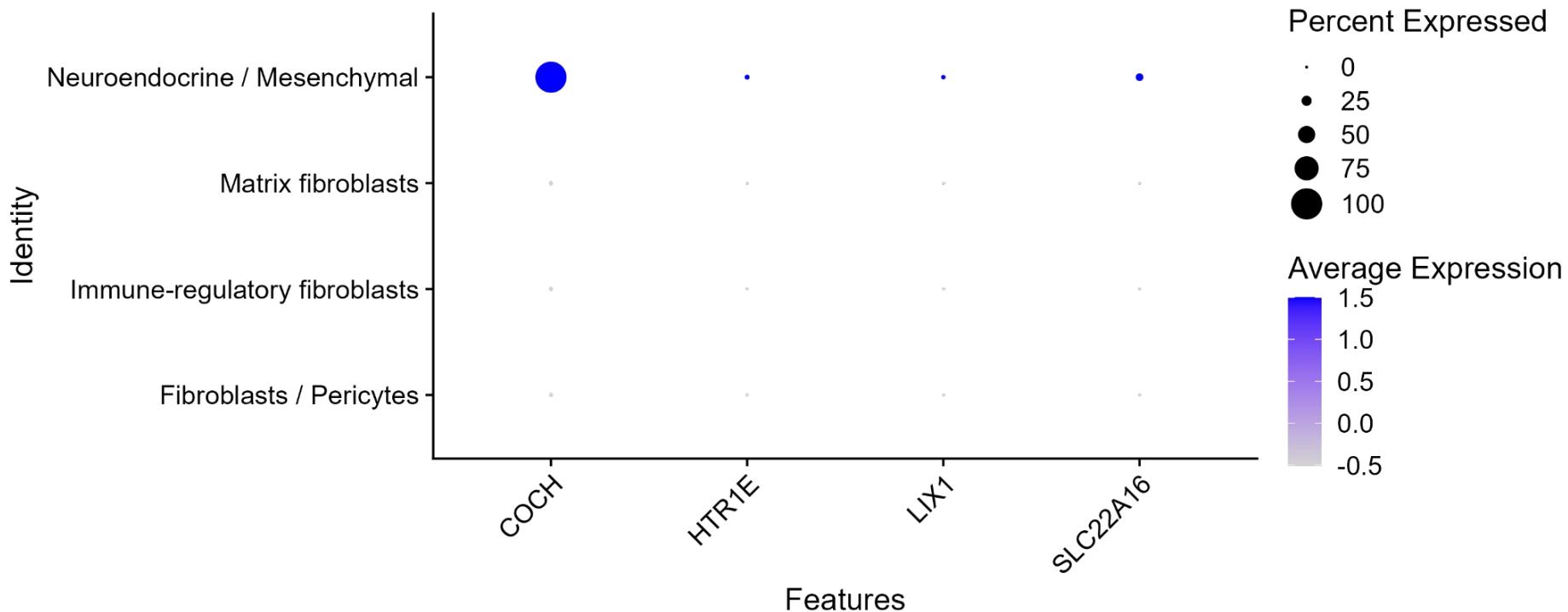


Блок 1.GO: Biological Process for: Neuroendocrine / Mesenchymal.



Топ-маркерні гени визначають гетерогенність фібробластів.

Neuroendocrine / Mesenchymal



Блок 2.Аналіз вже створеного SeuratObject зі статті.

Data: використовуючи Seurat::ReadRDS(), було прочитано SeuratObject. Об'єкт вже містив інтегровані датасети, а також інформацію про донорів (2 Young та 3 Old).

```
6_YOUNG 5_YOUNG 0_YOUNG 15_YOUNG 13_YOUNG 16_YOUNG 1_YOUNG 7_YOUNG 10_YOUNG  
1054     130      596      20       59      25      399      162      310  
8_YOUNG 2_YOUNG 3_YOUNG 9_YOUNG 4_YOUNG 14_YOUNG 12_YOUNG 11_YOUNG 6_OLD  
498      592      454      347      437      34      65      272      227  
15_OLD 5_OLD 7_OLD 13_OLD 0_OLD 16_OLD 1_OLD 3_OLD 2_OLD  
98      1269     907      168      1334     46      1487     1221     1201  
10_OLD 8_OLD 12_OLD 4_OLD 9_OLD 14_OLD 11_OLD  
188      224      229      1030     247      89      38
```

Кількість кластерів

```
[1] "0"  "1"  "2"  "3"  "4"  "5"  "6"  "7"  "8"  "9"  "10" "11" "12" "13" "14" "15"  
[17] "16"
```

Видаляємо всі інтеграційні результати – залишаємо тільки raw RNA-дані, щоб почати обробку з нуля.

Блок 2. Проводимо стандартний протокол для нового-старого Seurat

```
rna_obj_clean <- CreateSeuratObject(counts = seurat_art[["RNA"]])@counts)
```

```
rna_obj_clean <- NormalizeData(rna_obj_clean)
```

```
rna_obj_clean <- FindVariableFeatures(rna_obj_clean)
```

```
rna_obj_clean <- ScaleData(rna_obj_clean)
```

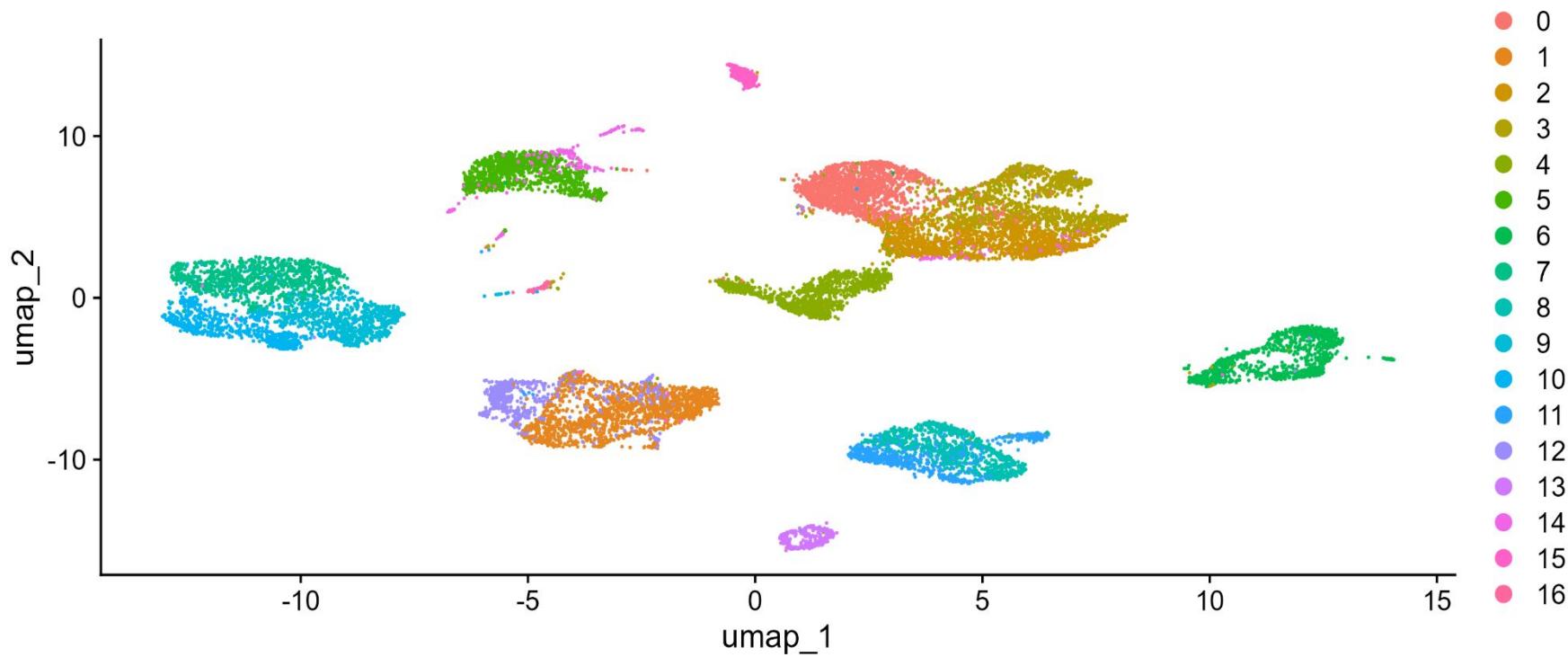
```
rna_obj_clean <- RunPCA(rna_obj_clean)
```

```
rna_obj_clean <- FindNeighbors(rna_obj_clean)
```

```
rna_obj_clean <- FindClusters(rna_obj_clean, resolution = 0.5)
```

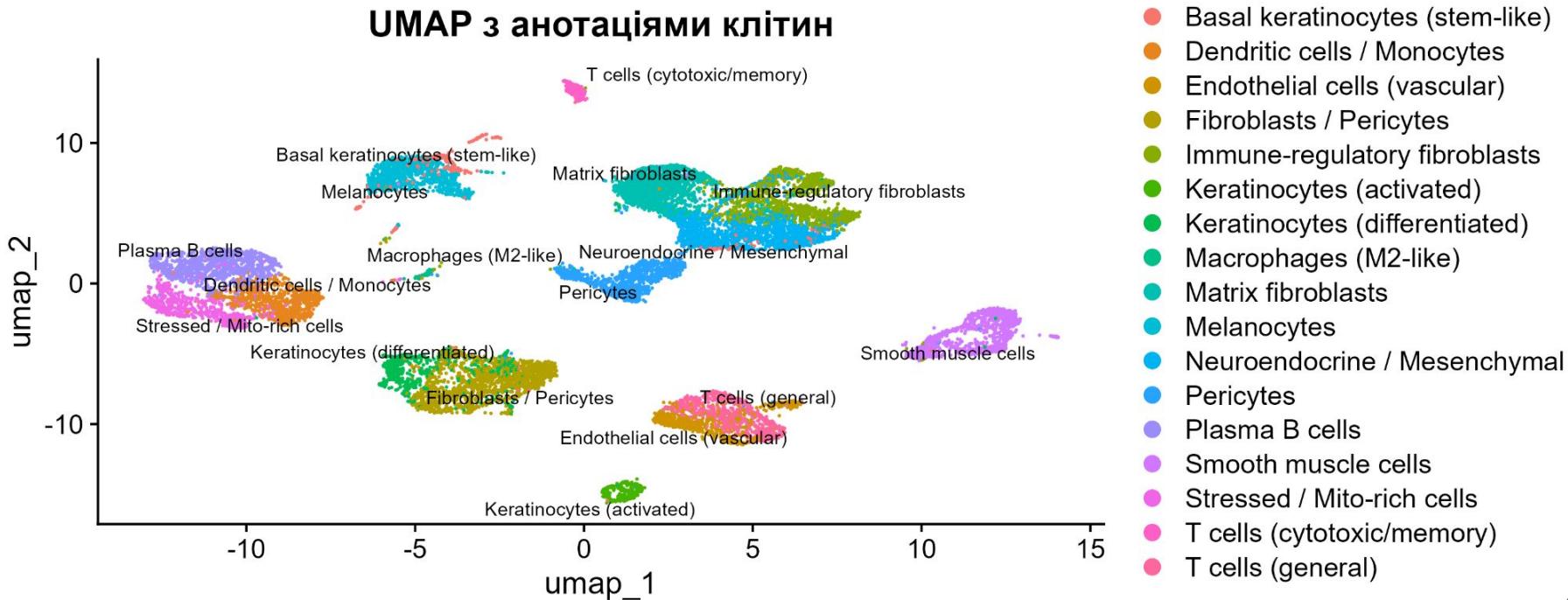
```
rna_obj_clean <- RunUMAP(rna_obj_clean, dims = 1:20)
```

Блок 2. UMAP

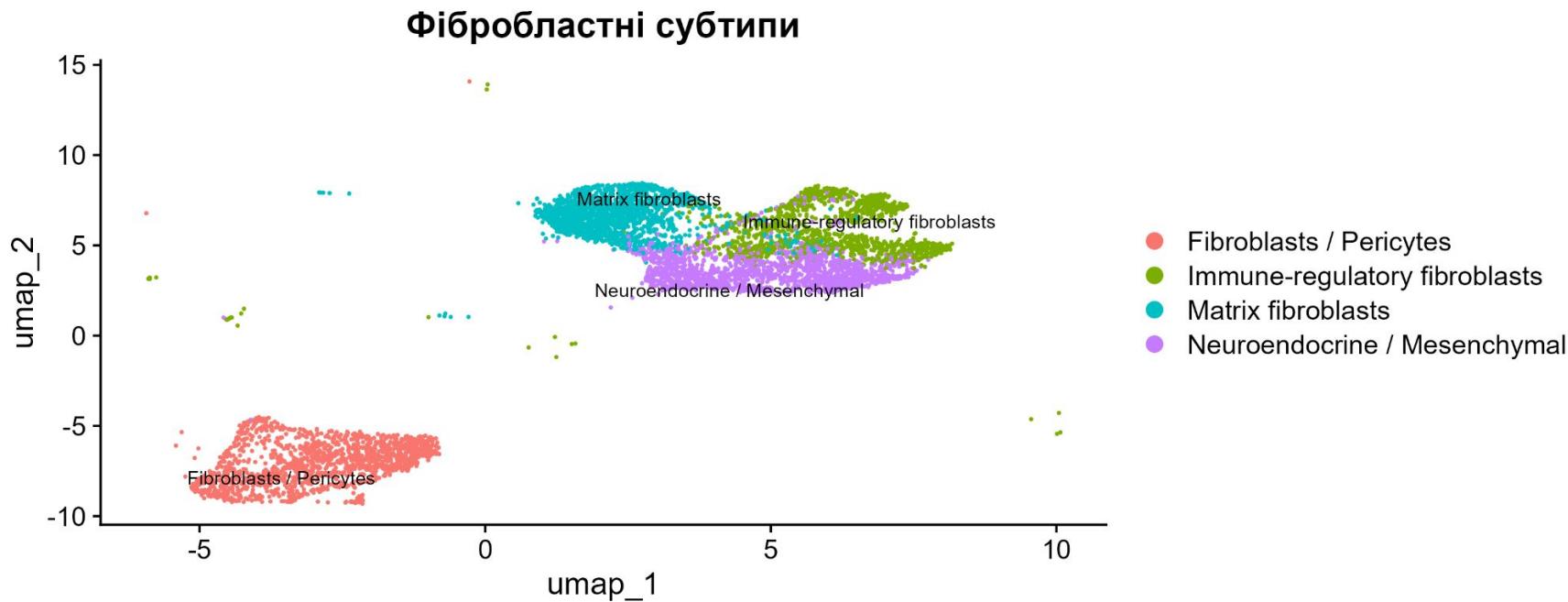


Блок 2. УМАР з анотаціями клітин.

Відповідно маркерам з блоку 1. (пошук за маркерними генами - топ 100)



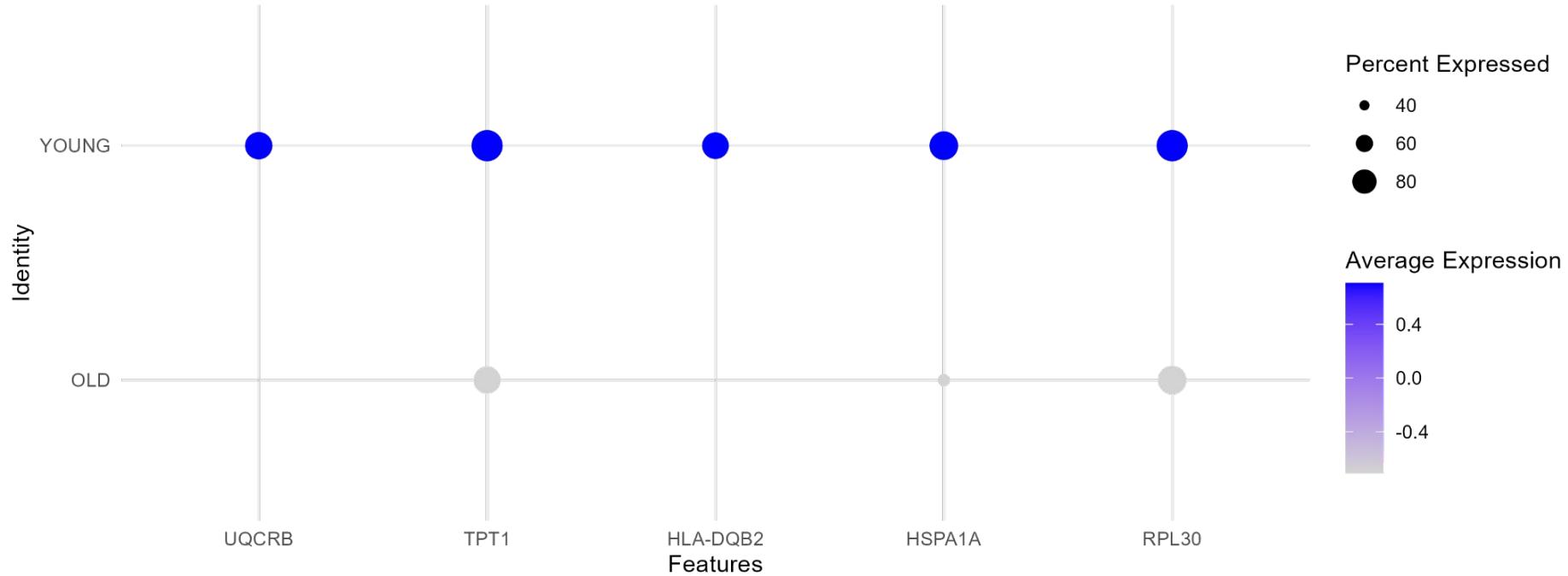
Блок 2. УМАР фібробластних субтипів.



Блок 2. OLD vs YOUNG

Fibroblasts / Pericytes.

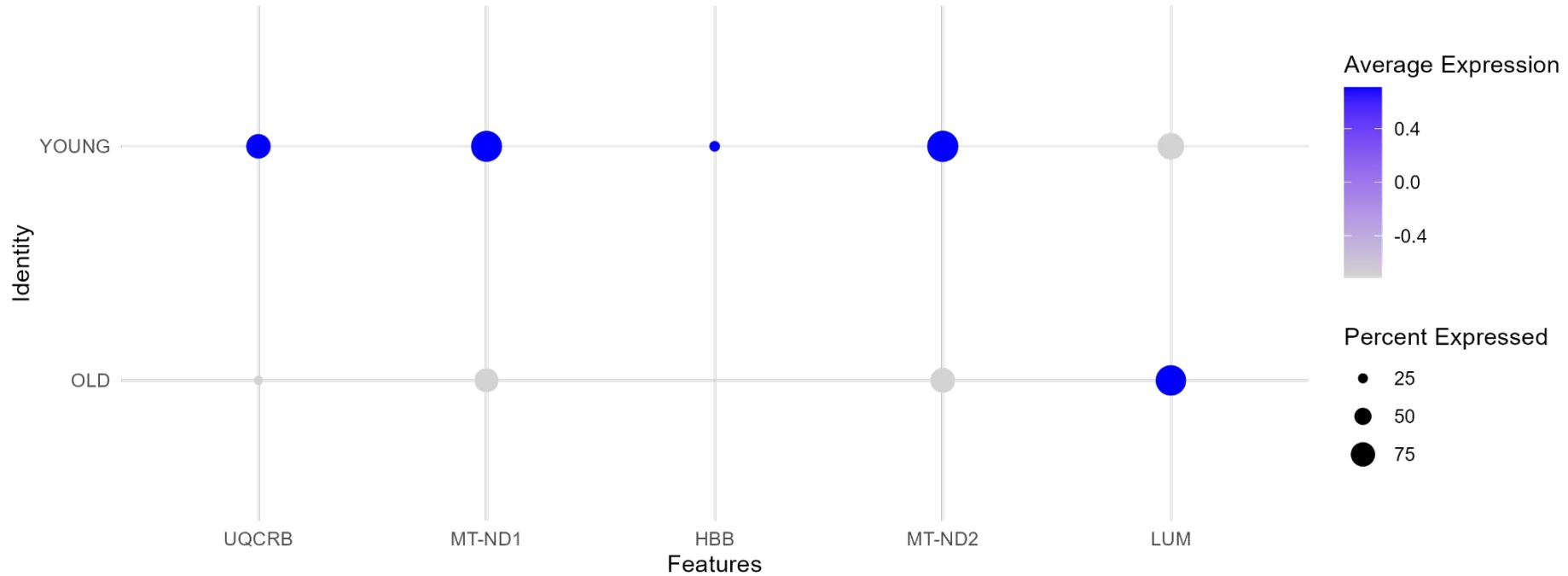
Top 5: OLD vs YOUNG – Fibroblasts / Pericytes



Блок 2. OLD vs YOUNG

Matrix fibroblasts.

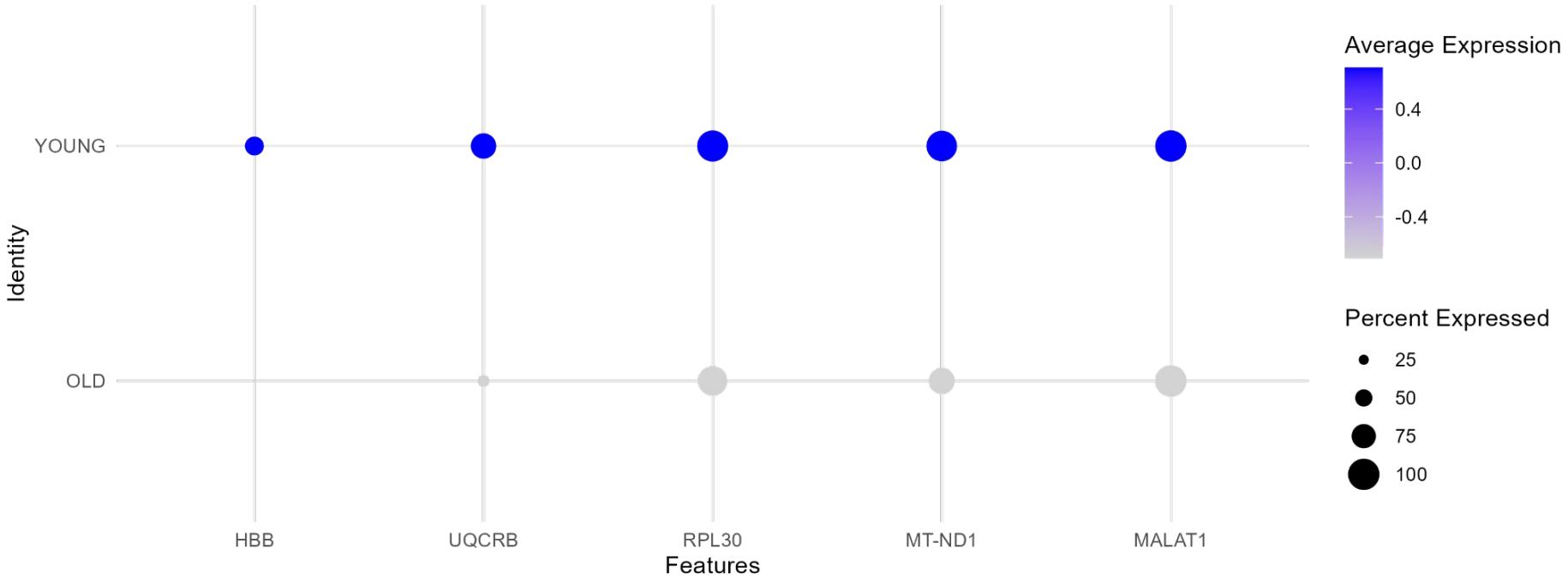
Top 5: OLD vs YOUNG – Matrix fibroblasts



Блок 2. OLD vs YOUNG

Immune-regulatory fibroblasts.

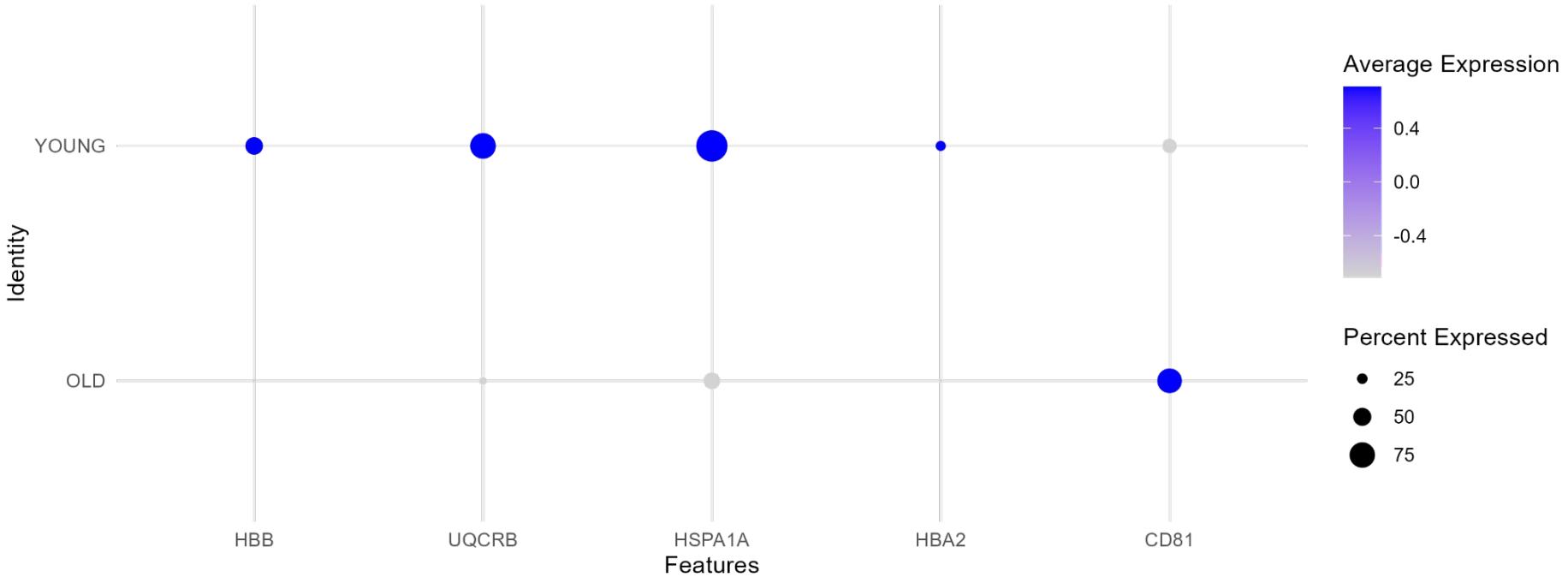
Top 5: OLD vs YOUNG – Immune-regulatory fibroblasts



Блок 2. OLD vs YOUNG

Neuroendocrine / Mesenchymal.

Top 5: OLD vs YOUNG – Neuroendocrine / Mesenchymal



Висновки.

1. Ідентифіковано чотири субтипи дермальних фібробластів, кожен із яких має унікальні функціональні характеристики згідно з GO-аналізом:
 - **Fibroblasts / Pericytes**: пов'язані з екстрацелюлярним матриксом, підтримкою епідермісу та регенерацією шкіри. Також відповідають за взаємодією з іншими шарами.
 - **Matrix fibroblasts**: більш виражено пов'язані з підтримкою організації дермального шару + на відміну від Fibroblasts / Pericytes мають менше різноманіття функцій.
 - **Immune-regulatory fibroblasts**: мають імунорегуляторні функції.
 - **Neuroendocrine / Mesenchymal**: пов'язані з відновленням тканини. Також активно беруть участь у розвитку клітин і обміні речовин. Мають нейрорегенеративний потенціал.

Висновки.

2. GO-аналіз, а також порівняння експресії маркерних генів між субтипами, підтвердило чітку функціональну диференціацію між підтипами, що свідчить про гетерогенність фібробластів у шкірі.
3. З віком фібробласти втрачають функціональну ідентичність – спеціалізація субтипів знижується.
4. Методологічні відмінності між Блоками 1 та 2 привели до різної кількості кластерів та варіацій у UMAP-візуалізації, проте ключові сигнатури збігаються (хоча для коректного порівняння знадобилась більша кількість маркерних генів).

Розподіл задач.

Oleksandr Surma - Блок 1 + доопрацювання Блоку 2 + презентація.

Anastasiiia Bliuma - робота з літературою та обробка Блоку 2.

Nikita Oleinikov - робота з літературою та анотація кластерів.

Використання генеративних моделей.



У процесі роботи над проєктом використовувався ChatGPT для траблшутінгу технічних помилок та загалом покращення розуміння мови програмування та методів аналізу scRNA-seq.

Це дозволило ефективніше вирішувати проблеми та глибше зануритись у біоінформатичні підходи, що використовувались у дослідженні.

ДЯКУЮ ЗА УВАГУ!!!