

Uniwersytet Jagielloński w Krakowie
Wydział Biochemii, Biofizyki i Biotechnologii

Załącznik do pracy licencjackiej: badanie wpływu G-CSF na transkryptom neutrofilii

Praca licencjacka na kierunku bioinformatyka

Miejsce wykonania pracy: Zakład Immunologii

Kraków, 2025

1 Pobieranie danych:

Do badania wykorzystano próbki pacjentów 1, 2 (grupa kontrolna) i 7, 8 (grupa badana), które zostały udostępnione w ramach pracy "Cellular and Transcriptional Dynamics of Human Neutrophils at Steady State and upon Stress" - [2] Pliki fastq pacjentów zostały udostępnione w bazie danych ArrayExpress pod numerem E-MTAB-11188. Pobrane pliki pochodziły z kategorii "submitted file"

Następnie pobrane pliki przesłano na platformę Galaxy (usegalaxy.eu) [1]

2 Przebieg analizy

1. Przetwarzanie danych wejściowych na platformie Galaxy (ETAP 1)
2. Dodanie metadanych przy użyciu języka programowania Python (ETAP2)
3. Dalsza analiza danych za pośrednictwem platformy Galaxy (ETAP 3)
4. Wizualizacja danych dotyczących genów markerowych z wykorzystaniem języków programowania Python i R (ETAP 4)
5. Wykonanie analizy wzbogacenia za pośrednictwem narzędzia DAVID - wykorzystanie języków programowania Python i R (ETAP 5)

Szczegółowe parametry narzędzi wykorzystanych do analizy na platformie Galaxy

ETAP 1 - Przetwarzanie danych wejściowych na platformie Galaxy

Na każdej z 4 próbek użyto narzędzi RNA STARsolo, DropletUtils i wygenerowano matrycę zliczeń

2.1 RNA STARSolo

Input Parameter	Value
Custom or built-in reference genome	indexed
Reference genome with annotation	without-gtf-with-gtf
Select reference genome	hg19full
Gene model file for splice junctions	Homo_sapiens.GRCh37.75.gtf
Elements from gene model to use	exon
Length around annotated junctions	100
Type of single-cell RNA-seq	CB_UMI_Simple
Input Type	repeat
Barcode reads file	Sample_scRNA_6_R1_001.fastq.gz
cDNA reads file	Sample_scRNA_6_R2_001.fastq.gz
Cell Barcode Whitelist	3M-february-2018.txt.gz
Configure Chemistry Options	Cv3
Barcode same size as Read	true
UMI deduplication algorithm	1MM_CR
Type of UMI filtering	Remove UMIs with N and homopolymers
Barcode matching strategy	Multiple matches (CellRanger 2, 1MM_multi)
Mode	solo
Library strandedness	Read strand same as original RNA
Genomic features for UMI counts	Gene: Count reads matching transcript
WASP filtering	Activated
Cell filtering type	no_filter
Output raw counts	true
Field 3 in gene output	Gene Expression
BAM alignment tags	NH, HI, AS, nM
Global gene count	false
Output unmapped reads	false
MAPQ for unique mappers	60
Max junctions per read	1000
Max collapsed junctions	1,000,000
Max genome inserts on the fly	1,000,000
Compute coverage	None
Wiggle output strand info	false

2.2 Usuwanie pustych kropelek (dropletów) DropletUtils

Input Parameter	Value
Format for the input matrix	directory
Count Data	RNA STARSolo on data 11, data 16, and others: Matrix Gene Counts raw
Genes List	RNA STARSolo on data 11, data 16, and others: Genes raw
Barcodes List	RNA STARSolo on data 11, data 16, and others: Barcodes raw
Operation	filter
Method	defaultdrops
Expected Number of Cells	10,000
Upper Quantile	0.99
Lower Proportion	0.1
Format for output matrices	Bundled (barcodes.tsv, genes.tsv, matrix.mtx)
Random Seed	100

2.3 Wygenerowanie matrycy zliczeń

Input Parameter	Value
Expression matrix (.mtx)	DropletUtils 10X Matrices on data 67, 66, 68
Gene table	DropletUtils 10X Genes on data 67, 66, 68
Barcode/cell table	RNA STARSolo on data 11, 16, others: Barcodes raw
Cell metadata table	—
Gene metadata table	—
Format of output object	AnnData format
Annotation index attribute	Gene ID
Output	Sample 6: anndata_h5ad AnnData

2.4 Scalanie obiektów Anndata

Input Parameter	Value
Annotated data matrix	Sample 1: anndata_h5ad AnnData
Function to manipulate the object	concatenate
Annotated data matrices to add	Sample 6, Sample 7, Sample 2: anndata_h5ad AnnData
Join method	Intersection of variables
Key for batch annotation in obs	patient
Merging strategy for uns	Default (result: empty dict)
Separator for index names and batch category	—
Output	Anndata_concatenate