



EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



Supported by:

**DAAD**  
Deutscher Akademischer Austausch Dienst  
German Academic Exchange Service

Bundesministerium für  
Bildung und Forschung

# РНК-секвенування - аналіз даних в R. Заняття 3-4.

---

Геноміка ЮА

Сергій Науменко



Геноміка ЮА — неприбуткова громадська організація  
ареєстрована в Україні: №43178260 від 16.08.2019

EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



**DAAD**  
Deutscher Akademischer Austausch Dienst  
German Academic Exchange Service



Bundesministerium für  
Bildung und Forschung

Альтернативна назва:

Аналіз даних РНК-секвенування: від транскриптів до  
біологічних інтерпретацій"

RNA-seq analysis: from count matrix to biological  
insights

# Організаційні питання

- Пишіть в головний чат, а не в чат зустрічі
- Словник:  
<https://github.com/GenomicsUA/awesome-ukrainian-omics-words>
- Розподіл на команди (<= 5 учасників) - до 12 серпня:  
<https://docs.google.com/spreadsheets/d/1UBIxwQW2LKFWFQJqiKzIqLD3RJsFefxhOrM77GHcd2g/edit?usp=sharing>
- Стандартний датасет і альтернативні датасети - обрати до 16.08
- Тести - з 9 по 23 вересня

# Основний датасет Petrenko2024

- Стаття: **Transcriptomic signatures of progressive and regressive liver fibrosis and portal hypertension**
- <https://pubmed.ncbi.nlm.nih.gov/38469563/>
- Матриця експресії, метадані:  
<https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-13804>
- Код у R:
  - <https://github.com/GenomicsUA/2024-daad-rnaseq-course>
- Практика: розбираємось із датасетом
- Домашнє завдання: повністю засвоїти датасет, приготувати питання

# Наш мінікурс – це базовий курс!

- Найпоширеніший аналіз
- Основи R
- Аналіз диференційної експресії (ДЕГ-аналіз)
- Основи DESeq2
- Короткі ріди (Illumina)
- Досить простий експеримент

# Критерії якості наукового проєкту для біоінформатика

- Наукова (біологічна, медична) проблема, експеримент
- Учасники проєкту
- Дані
- Методи

Чому варто  
досліджувати  
транскриптом?

# Чому варто досліджувати транскриптом?

- Інформація о біологічних функціях
- Експресія генів
- Регуляція експресії
- Біологія розвитку
- Біологія механізмів захворювань
- Транскрипти
- РНК-світ (походження біологічного життя)  
[https://en.wikipedia.org/wiki/List\\_of\\_RNAs](https://en.wikipedia.org/wiki/List_of_RNAs)
- РНК-віруси (Ebola, SARS, COVID-19, Influenza)  
Crick 1958, Crick 1970

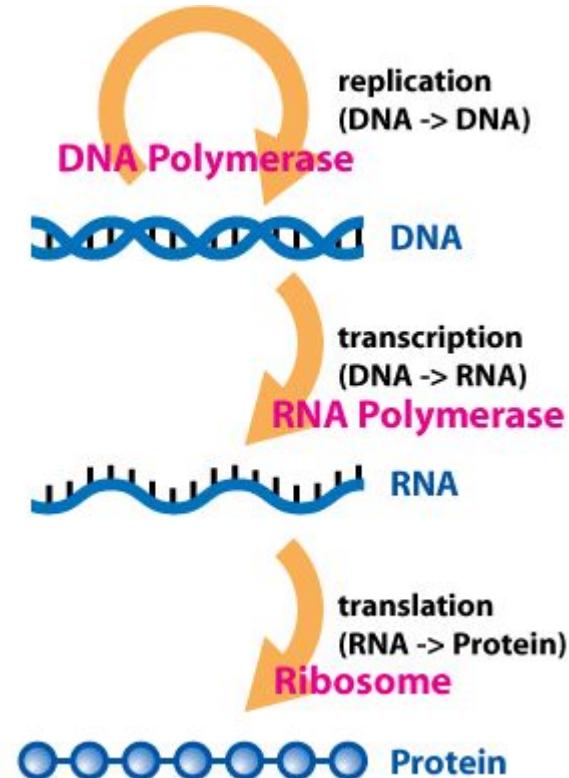


Рис - Вікіпедія

# Посилання

- Посібник з біоінформатики:  
[https://github.com/naumenko-sa/bioinf\\_posibnyk\\_public/blob/main/bioinf\\_posibnyk.pdf](https://github.com/naumenko-sa/bioinf_posibnyk_public/blob/main/bioinf_posibnyk.pdf)
- Kukurba and Montgomery. RNA sequencing and analysis. Cold Spring Harb Prot. 2015. Cited by 1133. Open access
- Van den Berge et al. RNA Sequencing Data: Hitchhiker's Guide to Expression Analysis. Annual Review of Biomedical Data Science. 2019. Cited by 141. Open access.
- Conesa et al. A survey of best practices for RNA-seq data analysis. Genome Biology. 2016. Cited by 3119. Open access.
- Stark et al. RNA sequencing: the teenage years. Nat. Rev Gen. 2019. Cited by 1678. (paywall but pdf is easily searchable).

## Посилання-2

- Курс від Harvard Chan School of Public Health Bioinformatics Core:  
[https://github.com/hbctraining/Training-modules/tree/master/planning\\_successful\\_rnaseq](https://github.com/hbctraining/Training-modules/tree/master/planning_successful_rnaseq)  
[https://github.com/hbctraining/DGE\\_workshop\\_salmon/blob/master/lessons/04\\_DGE\\_DESeq2\\_analysis.md](https://github.com/hbctraining/DGE_workshop_salmon/blob/master/lessons/04_DGE_DESeq2_analysis.md)
- RNA-seq від bioinformatics.ca, версія 2024:  
[https://github.com/bioinformaticsdotca/RNA\\_2024/blob/main/RNA\\_2024\\_main.md](https://github.com/bioinformaticsdotca/RNA_2024/blob/main/RNA_2024_main.md)
- DESeq2 documentation:  
<https://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html>

# Які типи аналізу можливі із даними РНК-секвенування?

- **Аналіз диференційної експресії генів (DEG, ДЕГ-аналіз)**
- збірка транскриптомів *de-novo*, яка актуальна для немодельних видів;
- аналіз мутацій сплайсингу, також і в менделівських захворюваннях;
- аналіз експресії ізоформ генів;
- аналіз експресії окремих екзонів;
- визначення ДНК-варіантів;
- аналіз A>I редактування;
- аналіз gene fusions;
- аналіз алель-специфічної експресії.

# Розвиток технологій дослідження транскриптому

- Історія: qPCR, EST, SAGE, CAGE and microarrays 1990-2010, see in kukurba2015.

<https://academic.oup.com/bib/article/8/1/6/264831>

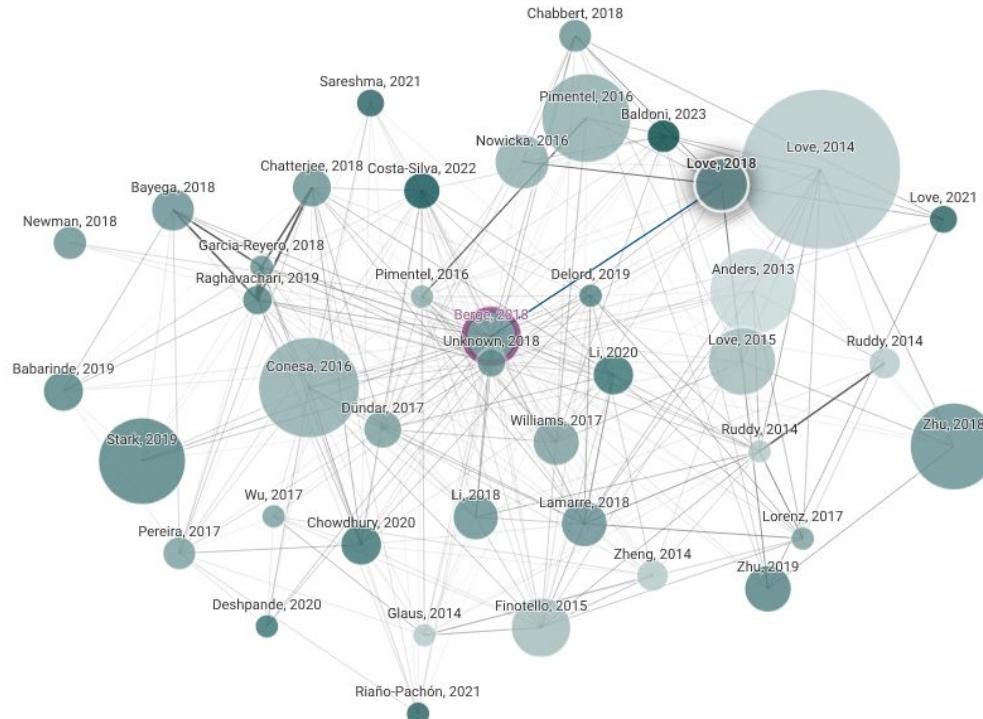
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2776927/>

- Wang2009. RNA-seq: a revolutionary tool for transcriptomics. 16K citations!

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2949280/>

- cDNA, short reads, long reads, direct sequencing - see fig1, table1 in Stark2019.

# Connected papers graph of RNA-seq analysis reviews



Key reviews are Stark2019, Berge 2018, Conesa 2016, note DEseq Love 2014

# Encode and Gencode projects

- <https://www.gencodegenes.org/>
- <https://www.encodeproject.org>

# GTEX (Genotype-Tissue expression project)

- <https://www.science.org/doi/10.1126/science.aaz1776>
- <https://gtexportal.org/home/publicationsPage>
- <https://gtexportal.org/home/>
- <https://gtexportal.org/home/tissueSummaryPage>:
- V8 release: 54 tissues, 948 donors, 17,382 samples
- Controlled access (via dbGAP)

Fig1. From Kukurba and Montgomery 2015

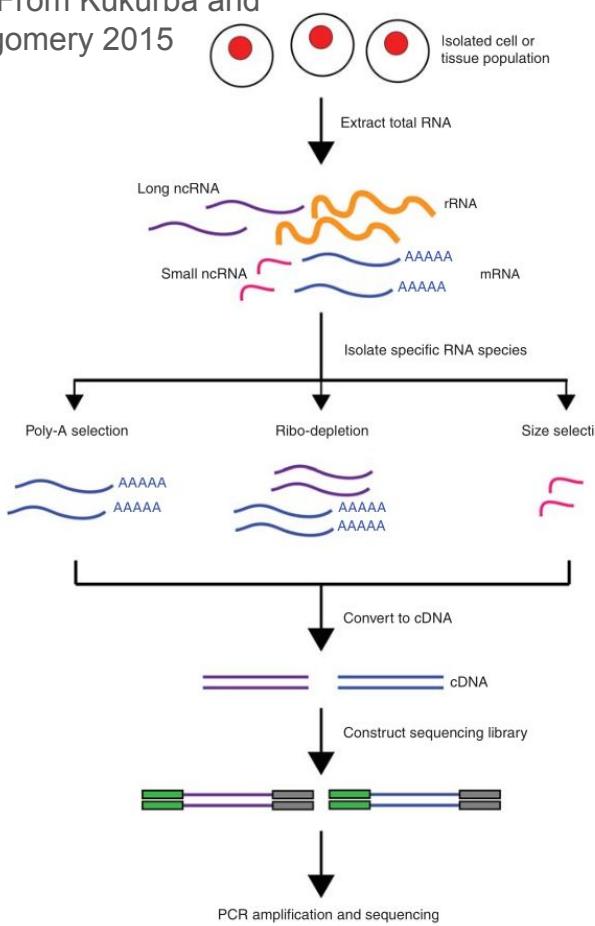


Fig1. from Van den Berge et al 2019

## Протокол РНК-сік експерименту

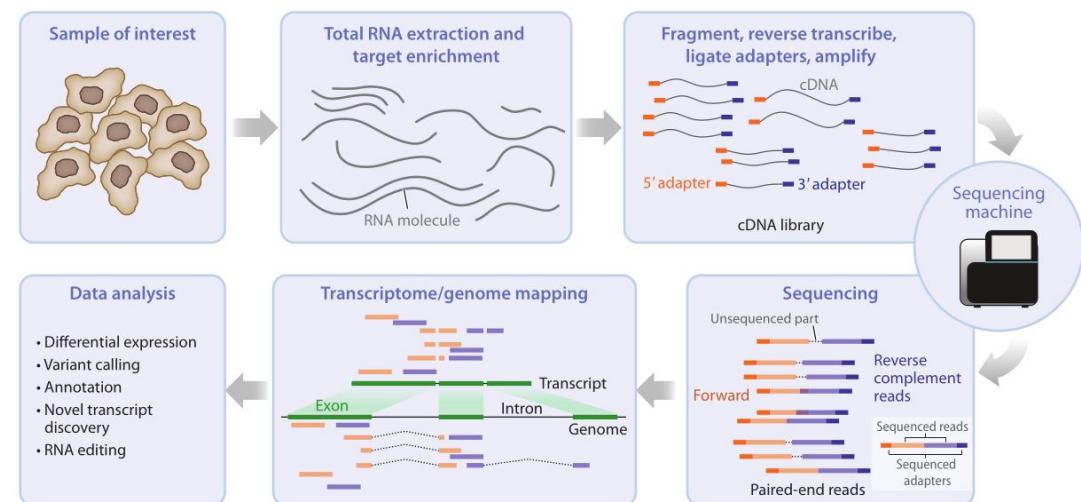


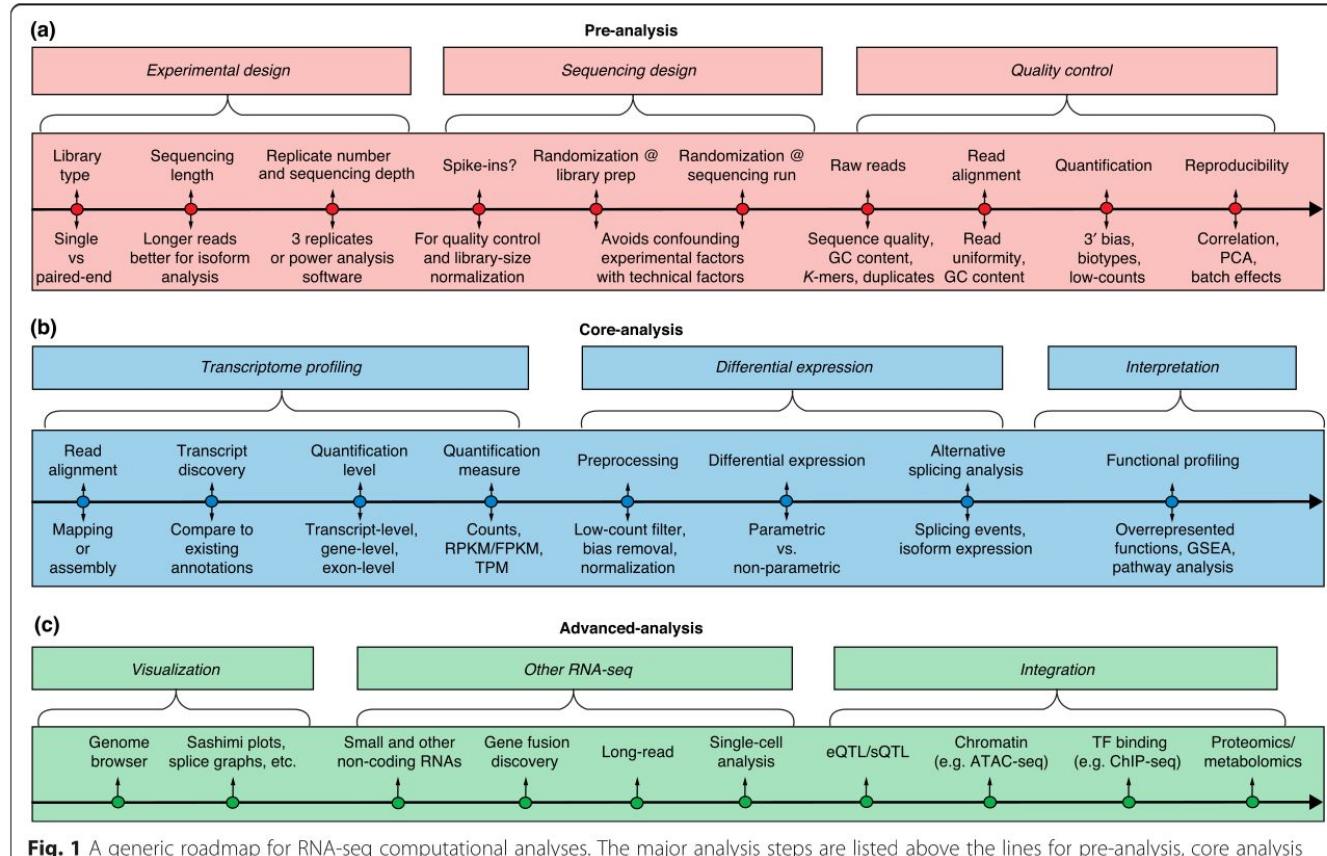
Figure 1

FIGURE 1. Overview of RNA-Seq. First, RNA is extracted from the biological material of choice (e.g., cells, tissues).

# Головні стадії ДЕГ-проєкту

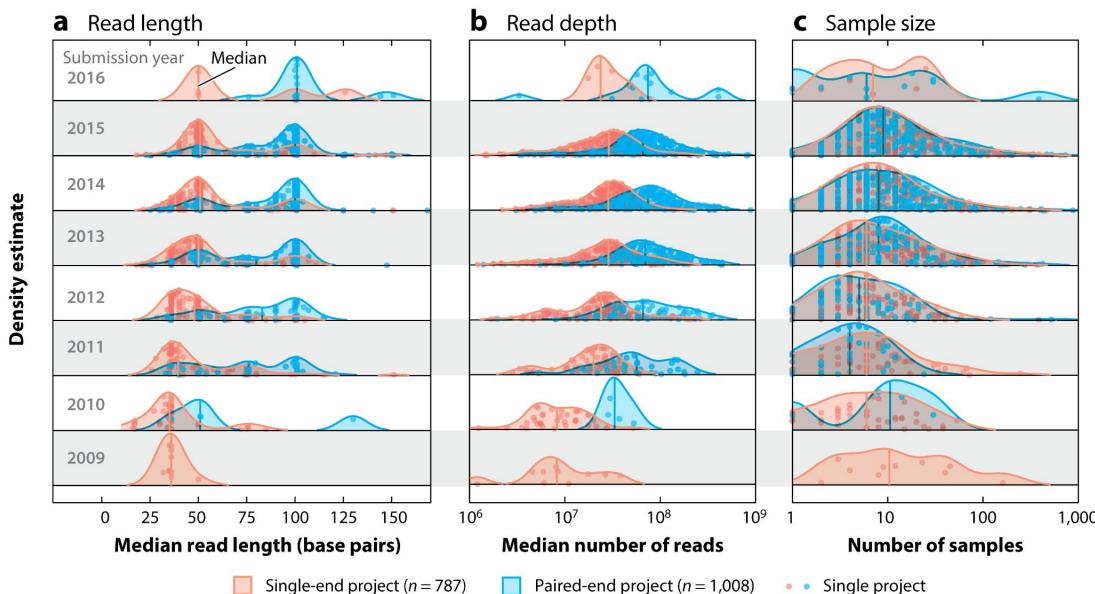
- планування експерименту — метадані, глибина секвенування, покриття (sequencing depth), репліки (replicates);
- контроль якості даних;
- отримання counts рівня генів чи транскриптів — матриця експресії
- аналіз диференційної експресії — набір up-regulated і down-regulated генів;
- функціональний аналіз груп генів.

# Шлях РНК-сік аналізу згідно Conesa et al 2016, fig.1



# Планування експерименту

- Скільки треба прочитань (reads)?
- Парні чи одинарні прочитання? ( $15 \text{ mln paired} = 30 \text{ mln single reads}$ )



# Планування експерименту

- За уроком від Harvard Chan School of Public Health bioinformatics core:  
[https://hbctraining.github.io/Intro-to-rnaseq-hpc-salmon/lessons/experimental\\_planning\\_considerations.html](https://hbctraining.github.io/Intro-to-rnaseq-hpc-salmon/lessons/experimental_planning_considerations.html)
- Репліки
- Confounding
- Batch effects
- Фактори!

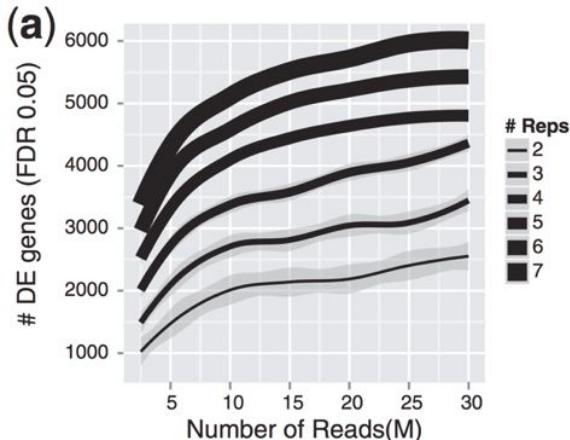
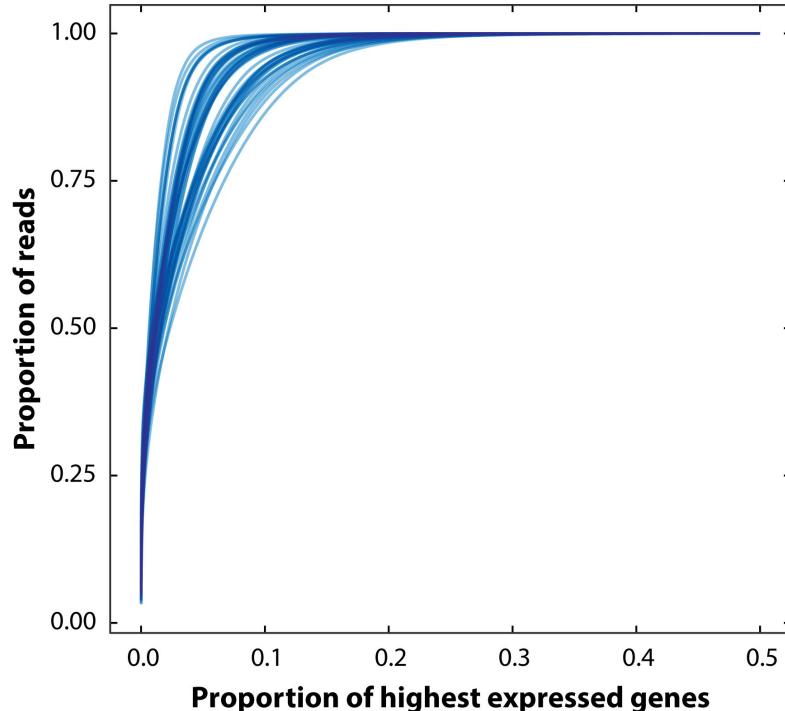


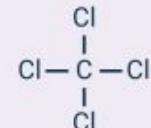
Image credit: [Liu, Y., et al., Bioinformatics \(2014\) 30\(3\): 301–304](#)



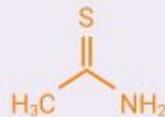
## Experimental design: induction and regression of liver fibrosis



Male C57BL/6JRj mice ( $n = 48$ )  
Fibrosis induction groups:

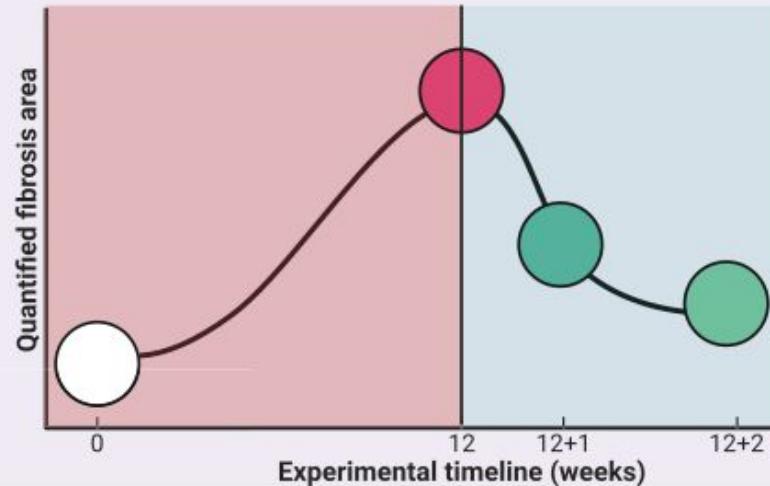


Carbon tetrachloride  
( $\text{CCl}_4$ ,  $n = 18$ ) or  
olive oil control



Thioacetamide  
(TAA,  $n = 18$ ) or  
saline control

Fibrosis induction phase      Fibrosis regression phase



Метадані див в petrenko2024.xlsx

# Планування експерименту

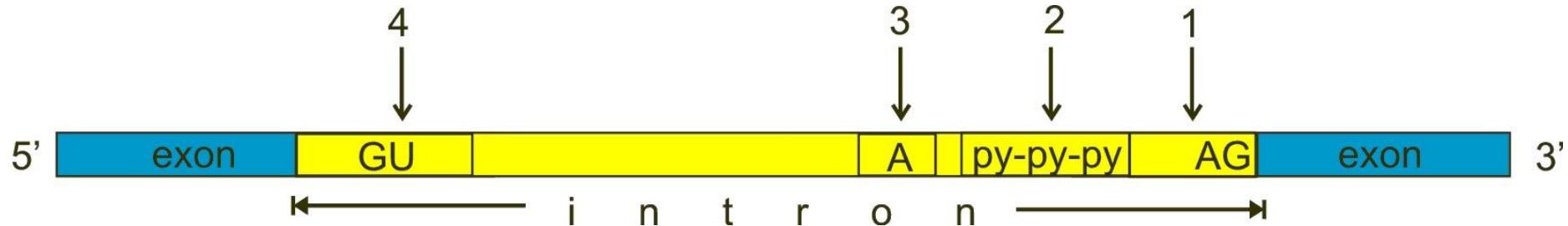
- RIN = RNA Integrity Number ( $>=7$ ), measured after RNA isolation
- RNA capture protocol (polyA, ribo-Depletion, capture Seq/pulldown)?
- Stranded, unstranded?
- spike-ins?

# Контроль якості даних

- Повинен бути реалізований в лабораторному пайплайні
- Fastqc, multiqc - якість секвенування
- % aligned reads
- % aligned to coding sequences, rRNA
- 5'-3' bias
- N reads
- N genes detected
- Gene detection saturation
- Порівняти зразки, чи є викиди (outliers)
- PCA - чи є сенс?
- Чи треба виключити якісь дані з аналізу?
- Приклад: див. quality control in Mair2020

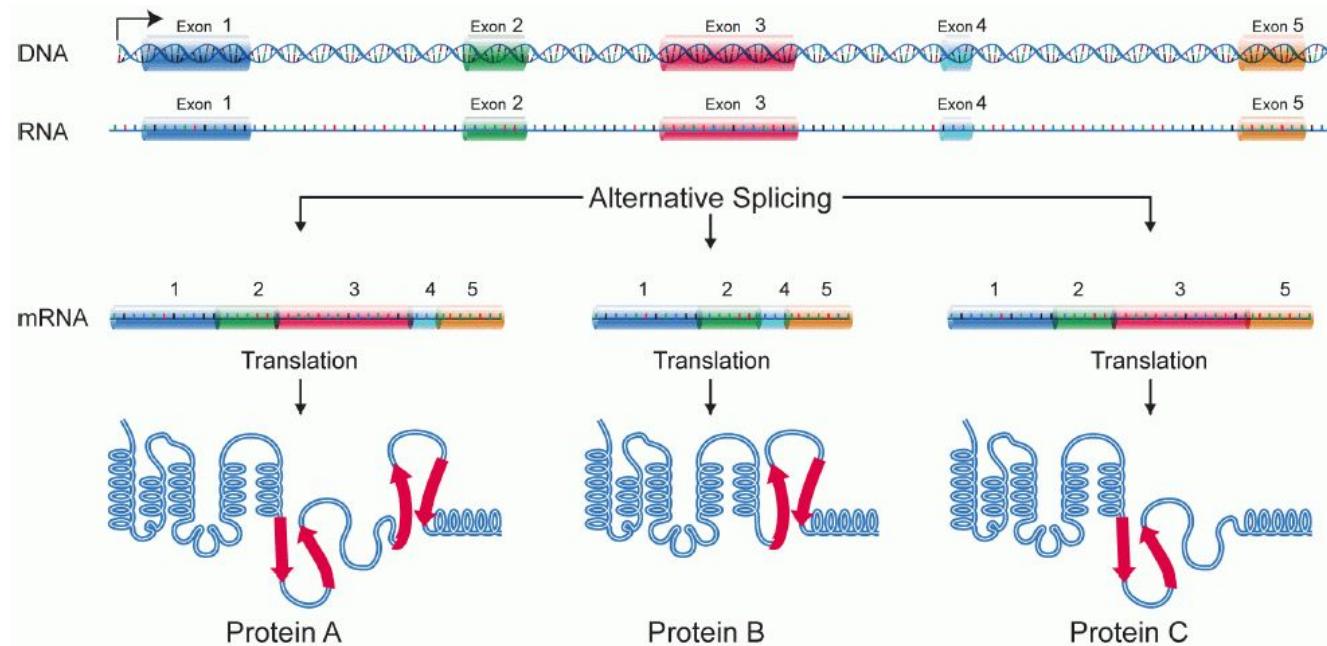
# Треба нагадати про сплайсинг?

- [https://en.wikipedia.org/wiki/RNA\\_splicing](https://en.wikipedia.org/wiki/RNA_splicing)



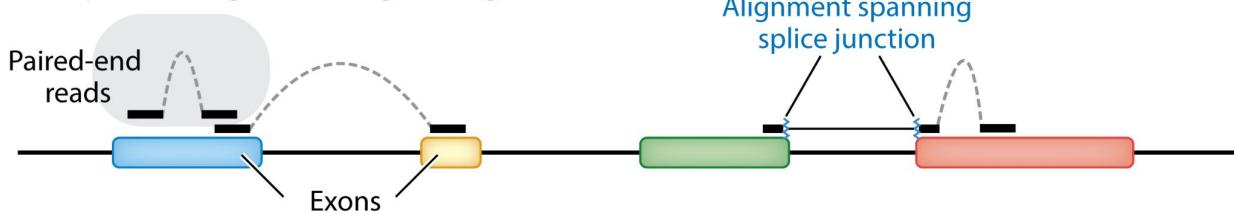
# А про альтернативний сплайсинг?

- [https://en.wikipedia.org/wiki/Alternative\\_splicing](https://en.wikipedia.org/wiki/Alternative_splicing)

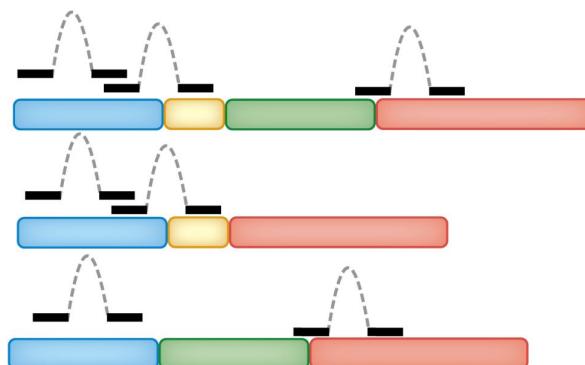


# Spliced alignment

**a** Spliced alignment against genome



**b** Unspliced alignment against transcriptome



Van den Berge K, et al. 2019.  
*Annu. Rev. Biomed. Data Sci.* 2:139–73

# Abundance estimation

- 1st generation: featureCounts
- 2nd generation: rsem (see next slide)
- 3rd generation: kmer-based pseudoalignment (kallisto, salmon)
- 4th generation: selective alignment.

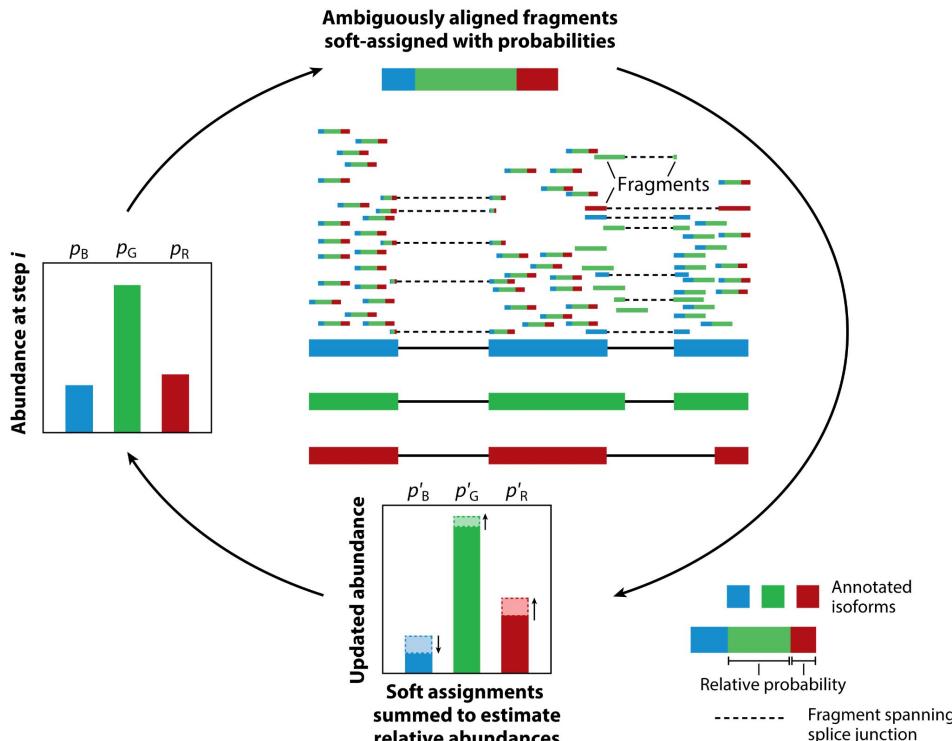
Salmon - selective alignment:

- <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02151-8>
- <https://salmon.readthedocs.io/en/latest/salmon.html>

Kallisto pipeline:

- <https://pachterlab.github.io/kallisto/>

# Схема RSEM - алгоритму



# На практиці

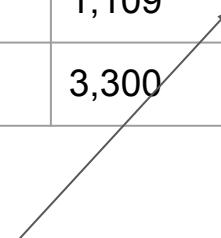
- STAR/hisat2
- <https://nf-co.re/rnaseq/3.14.0/>

Як виглядає файл із вирівнюванням (bam) - див. урок від НВС

[https://hbctraining.github.io/Intro-to-rnaseq-hpc-O2/lessons/04\\_alignme nt\\_quality.html](https://hbctraining.github.io/Intro-to-rnaseq-hpc-O2/lessons/04_alignme nt_quality.html)

# Матриця експресії

	Зразок1-е	Зразок2-е	Зразок3-е	Зразок4-к	Зразок5-б	Зразок6-к
Ген1	100	105	110	50	55	54
Ген2	1,000	1,109	900	1,200	1,000	800
Ген3	3,000	3,300	3,100	5,000	4,000	4,300



Значення експресії

Які гени (транскрипти, екзони) експресуються сильніше (слабкіше) у експерименті порівняно із контролем?

# Завдання додому: обрати цікавий датасет

- SARS-CoV-2 productively infects human gut enterocytes
  - <https://www.science.org/doi/full/10.1126/science.abc1669>
  - <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149312>
- <http://rna.recount.bio/>
- <https://bioconductor.org/packages/release/bioc/vignettes/recount3/inst/doc/recount3-quickstart.html>
- Candidate gene networks and blood biomarkers of methamphetamine-associated psychosis: an integrative RNA-sequencing report: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5070070/>
- [https://scholar.google.com/scholar?cites=16121678637925818947&as\\_sdt=2005&sciodt=0,5&hl=en](https://scholar.google.com/scholar?cites=16121678637925818947&as_sdt=2005&sciodt=0,5&hl=en)

# Нормалізація матриці експресії

- НВС:  
[https://hbctraining.github.io/DGE\\_workshop\\_salmon/lessons/02\\_DGE\\_count\\_normalization.html](https://hbctraining.github.io/DGE_workshop_salmon/lessons/02_DGE_count_normalization.html)
- StatQuest:
  - TPM/RPKM: <https://www.youtube.com/watch?v=TTUrtCY2k-w>
  - DESEQ2: <https://www.youtube.com/watch?v=UFB993xufUU>
- На глибину покриття
- На довжину генів
- На склад бібліотеки
- Більше детально - див. заняття 5-6

# Нормалізація – RPKM

The screenshot shows a YouTube video player with the URL <https://www.youtube.com/watch?v=TTUrTCY2k-w>. The video title is "RPKM Summary".

**BEFORE**

Gene Name	Rep1 Counts	Rep2 Counts	Rep3 Counts
A (2kb)	10	12	30
B (4kb)	20	25	60
C (1kb)	5	8	15
D (10kb)	0	0	1

Read counts were...

- 1) Normalized for differences in sequencing depth.
- 2) Normalized for gene size.

**AFTER**

Gene Name	Rep1 RPKM	Rep2 RPKM	Rep3 RPKM
A (2kb)	1.43	1.33	1.42
B (4kb)	1.43	1.39	1.42
C (1kb)	1.43	1.78	1.42
D (10kb)	0	0	0.009

RPKM, FPKM and TPM, Clearly Explained!!!

StatQuest with Josh Starmer 1.22M subscribers

Join    Subscribe

2.1K    Share    Thanks    Clip    Save

# Нормалізація – TPM

The screenshot shows a YouTube video player with the URL <https://www.youtube.com/watch?v=TTUrtCY2k-w>. The video title is "TPM – step 2: normalize for sequencing depth".

On the left, there is a vertical black bar with the text "TPM – scaled by gene length and sequencing depth (M):".

The main content area displays two tables:

**TPM – step 2: normalize for sequencing depth**

Gene Name	Rep1 RPK	Rep2 RPK	Rep3 RPK
A (2kb)	5	6	15
B (4kb)	5	6.25	15
C (1kb)	5	8	15
D (10kb)	0	0	0.1

Total RPK: 15      20.25      45.1  
Tens of RPK: 1.5      2.025      4.51

Gene Name	Rep1 TPM	Rep2 TPM	Rep3 TPM
A (2kb)	3.33	2.96	3.326
B (4kb)	3.33	3.09	3.326
C (1kb)	3.33	3.95	3.326
D (10kb)	0	0	0.02

At the bottom of the video player, the text "RPKM, FPKM and TPM, Clearly Explained!!!"



StatQuest with Josh Starmer

•

Join

Subscribe

2.1K

Share

Thanks

Clip

Save

...

# Нормалізація – RPKM vs TPM

The screenshot shows a YouTube video player with the title "RPKM vs TPM".

**RPKM Table:**

Gene Name	Rep1 RPKM	Rep2 RPKM	Rep3 RPKM
A (2kb)	1.43	1.33	1.42
B (4kb)	1.43	1.39	1.42
C (1kb)	1.43	1.78	1.42
D (10kb)	0	0	0.009
Total:	4.29	4.5	4.25

**TPM Table:**

Gene Name	Rep1 TPM	Rep2 TPM	Rep3 TPM
A (2kb)	3.33	2.96	3.326
B (4kb)	3.33	3.09	3.326
C (1kb)	3.33	3.95	3.326
D (10kb)	0	0	0.02
Total:	10	10	10

Annotations in the video player:

- A red oval highlights the "Total" row in the RPKM table.
- A red oval highlights the "Total" row in the TPM table.
- A text overlay on the left side of the RPKM table says "... the sums of each column are very different."

YouTube player controls at the bottom include: play/pause, volume, progress bar (7:50 / 10:14), and a search bar.

Video stats at the bottom: 2.1K likes, Share, Thanks, Clip, Save, and more.

Channel info: StatQuest with Josh Starmer, 1.22M subscribers.

Page number: 35

# Схема ДЕГ-аналізу згідно курсу від НВС

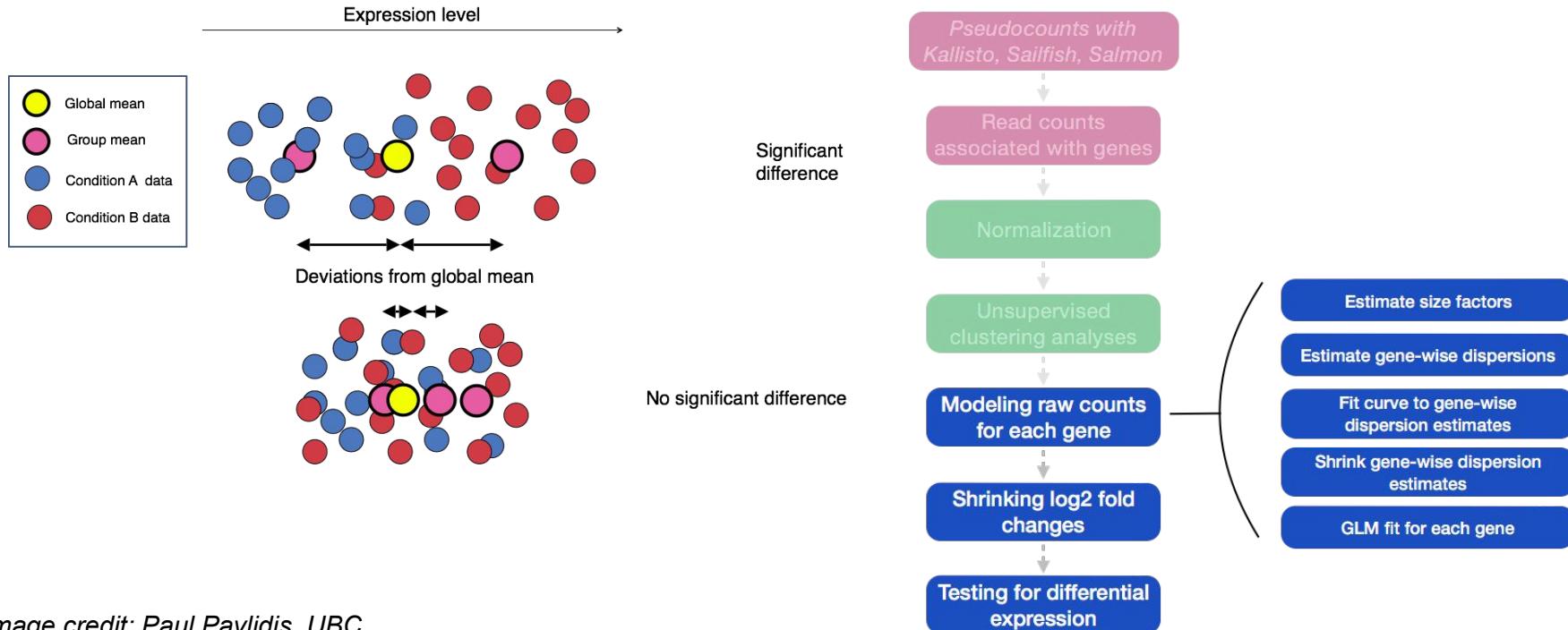
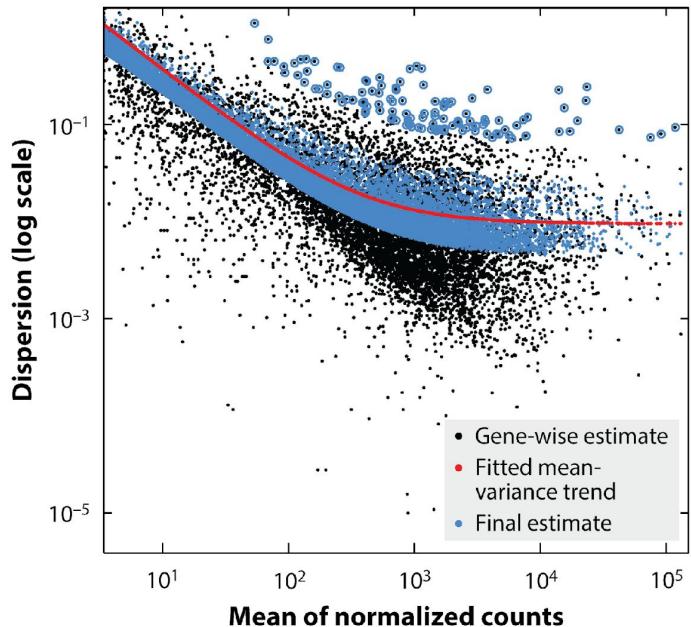
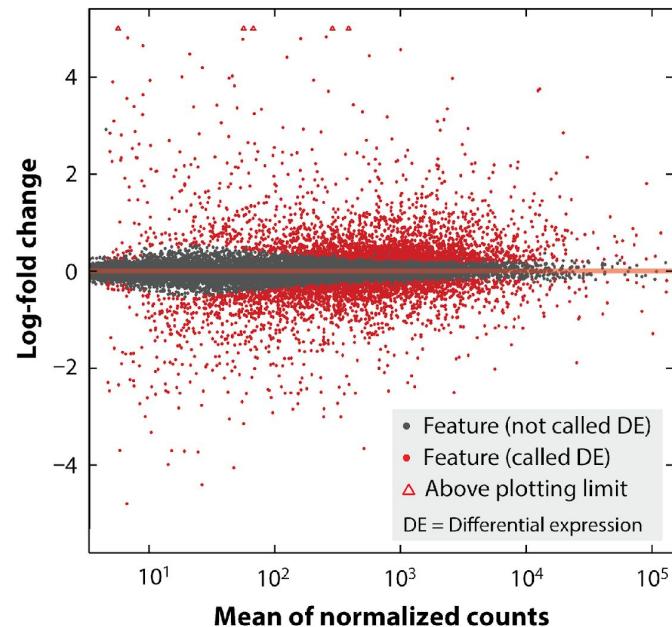


Image credit: Paul Pavlidis, UBC

[https://github.com/hbctraining/DGE\\_workshop\\_salmon/blob/master/lessons/04\\_DGE\\_DESeq2\\_analysis.md](https://github.com/hbctraining/DGE_workshop_salmon/blob/master/lessons/04_DGE_DESeq2_analysis.md)

**a****b**

Van den Berge K, et al. 2019.  
*Annu. Rev. Biomed. Data Sci.* 2:139–73

- Var > mean - using NB to model

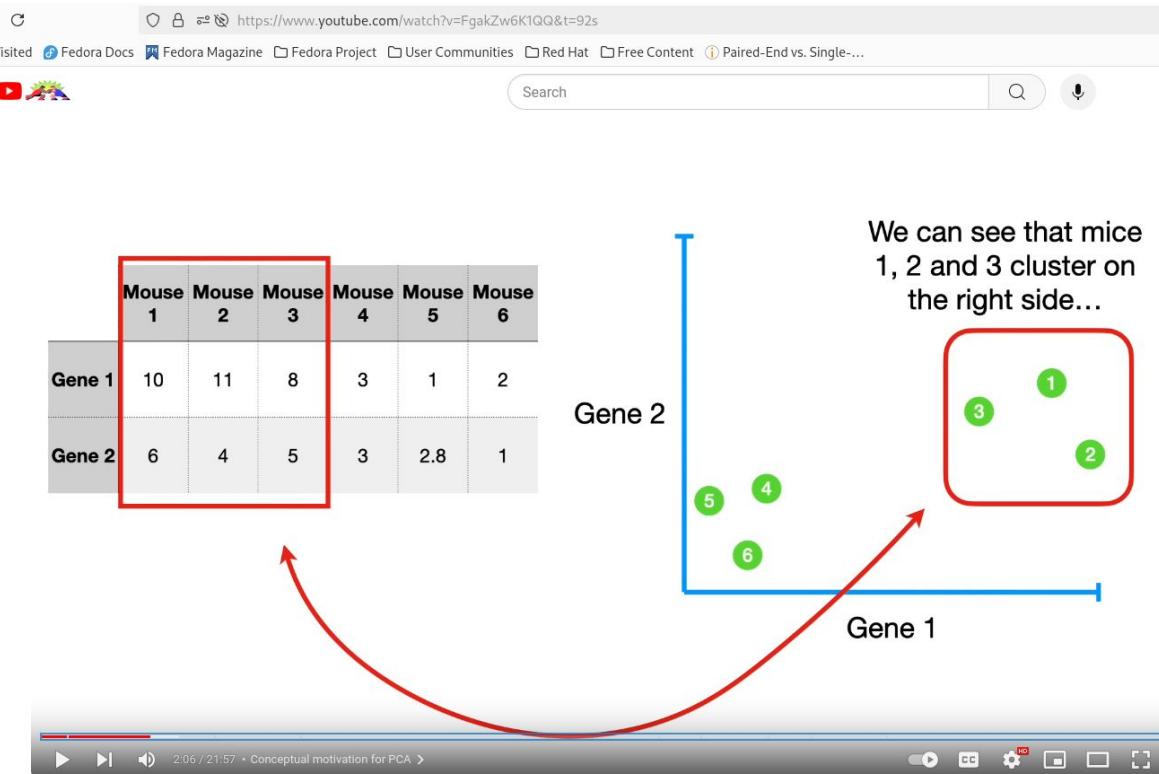
# Negative binomial distribution: Від'ємний біноміальний розподіл

- Біноміальний: моделюємо кількість успішних випробувань із ( $N$ ,  $p$ )
- ВБ: моделюємо кількість невдач в експерименті до  $r$  успіху ( $r$ ,  $p$ )  
<https://rpubs.com/mpfoley73/458738>
- [https://uk.wikipedia.org/wiki/Від'ємний\\_біноміальний\\_розподіл](https://uk.wikipedia.org/wiki/Від'ємний_біноміальний_розподіл)
- Просте пояснення, чому саме NB працює для РНК-сік даних:  
[https://www.youtube.com/watch?v=lJw6Ku\\_jQkM](https://www.youtube.com/watch?v=lJw6Ku_jQkM)
- [https://github.com/GenomicsUA/2024-daad-rnaseq-course/blob/main/02\\_classes/04\\_rna\\_seq\\_intro\\_part2/06.negative\\_binomial.Rmd](https://github.com/GenomicsUA/2024-daad-rnaseq-course/blob/main/02_classes/04_rna_seq_intro_part2/06.negative_binomial.Rmd)

# Терміни

- ◎ mRNA - матрична РНК
- ◎ Read, sequence read - прочитання

# PCA – метод головних компонент



StatQuest: Principal Component Analysis (PCA), Step-by-Step



StatQuest with Josh Starmer  
1.22M subscribers

Join

Subscribe



58K



...



Share



Thanks



Clip



Save



...

# PCA – метод головних компонент

https://www.youtube.com/watch?v=FgakZw6K1QQ&t=92s

/visited Fedora Docs Fedora Magazine Fedora Project User Communities Red Hat Free Content Paired-End vs. Single-...

Search

$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2 = \text{sum of squared distances} = \text{SS(distances)}$

...and we repeat until we end up with the line with the largest sum of squared distances between the projected points and the origin.

9:02 / 21:57 • Finding PC1 >

StatQuest: Principal Component Analysis (PCA), Step-by-Step



StatQuest with Josh Starmer  
1.22M subscribers

Join

Subscribe

58K

1

Share

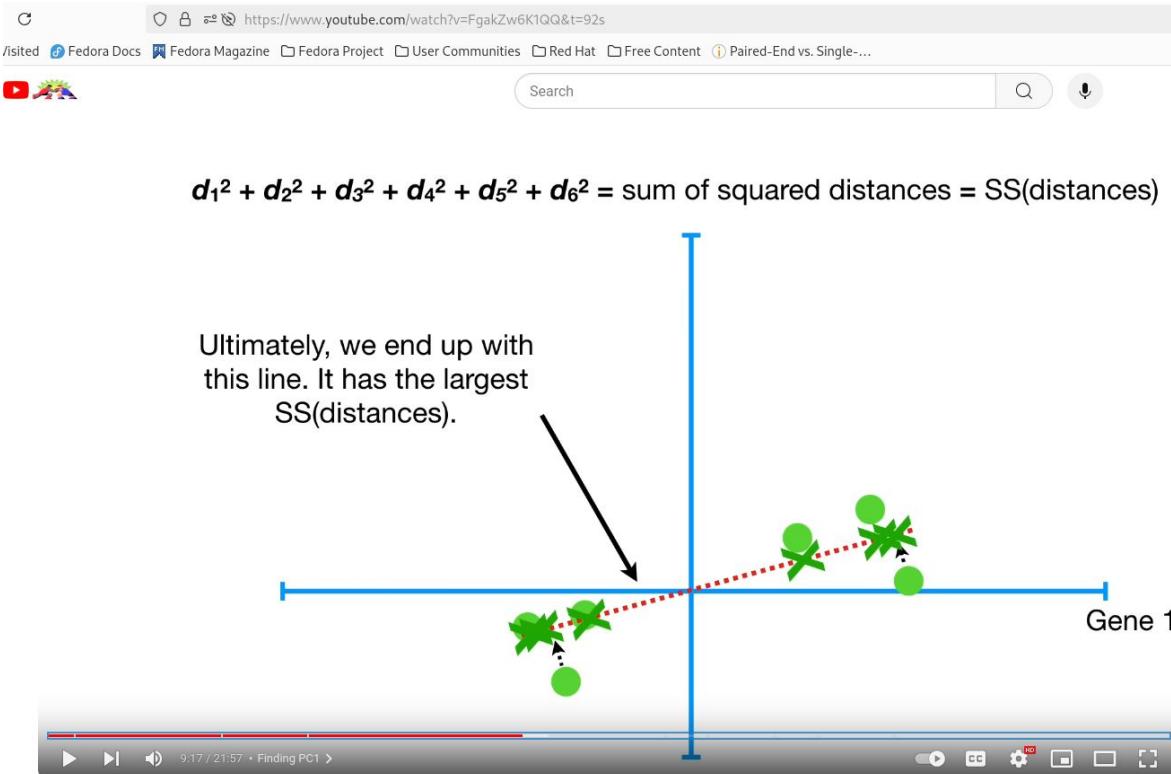
Thanks

Clip

Save

...

# PCA – метод головних компонент



StatQuest with Josh Starmer  
Join [Subscribe](#)

58K



Share

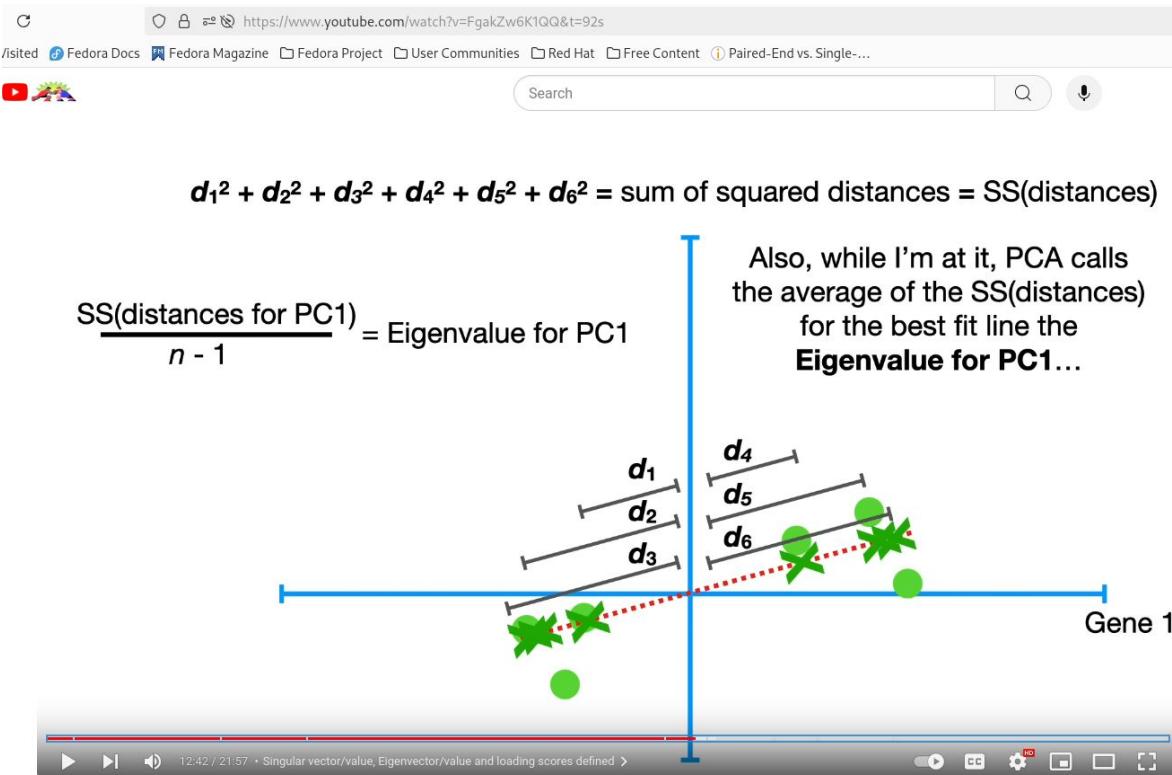


Clip



Save ...

# PCA – метод головних компонент



StatQuest: Principal Component Analysis (PCA), Step-by-Step



StatQuest with Josh Starmer

1.22M subscribers

Join

Subscribe

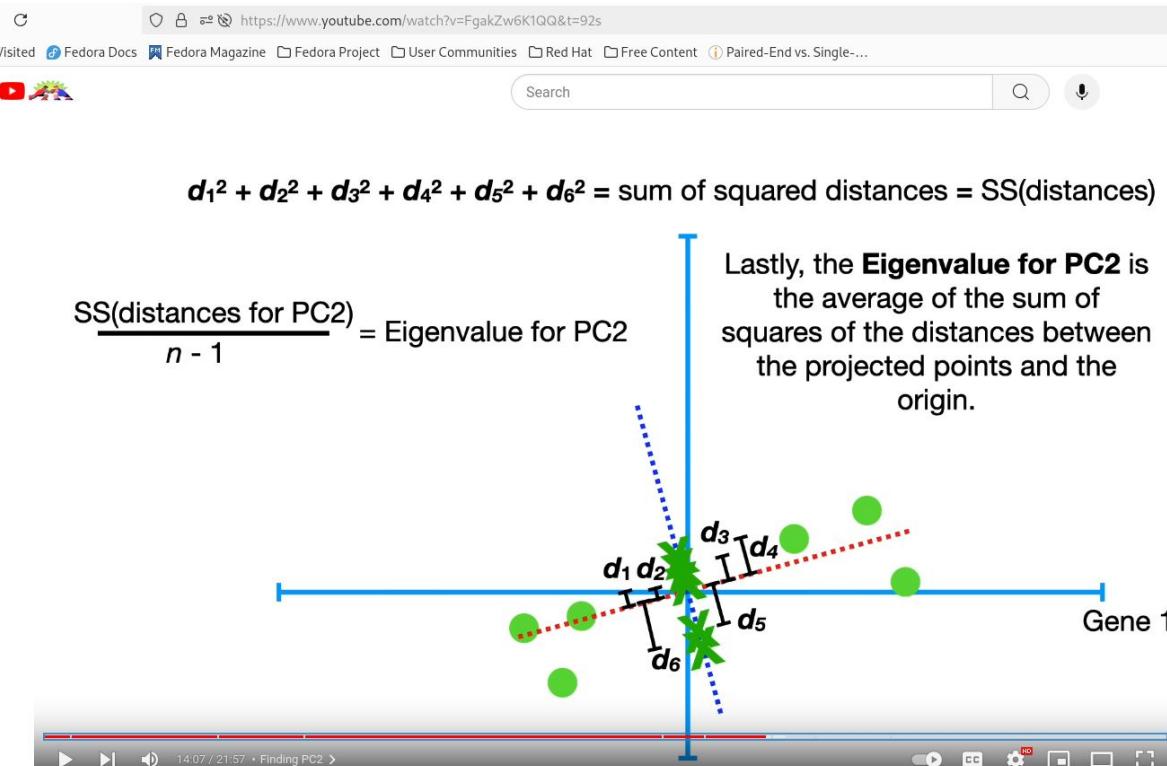
58K

Clip

Thanks

Save

# PCA – метод головних компонент



StatQuest: Principal Component Analysis (PCA), Step-by-Step



StatQuest with Josh Starmer  
1.22M subscribers

Join

Subscribe



58K



...



Share



Thanks



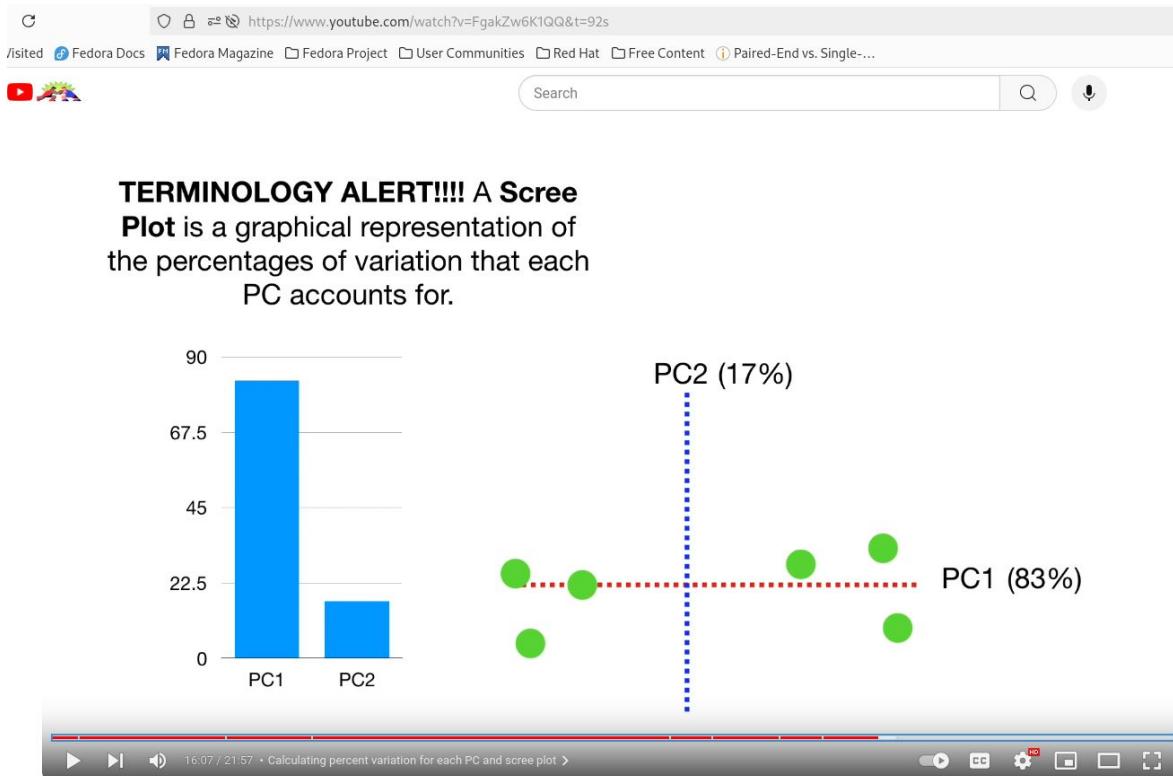
Clip



Save

...

# PCA – метод головних компонент



StatQuest: Principal Component Analysis (PCA), Step-by-Step



Join

Subscribe



58K



...



Share



Thanks



Clip



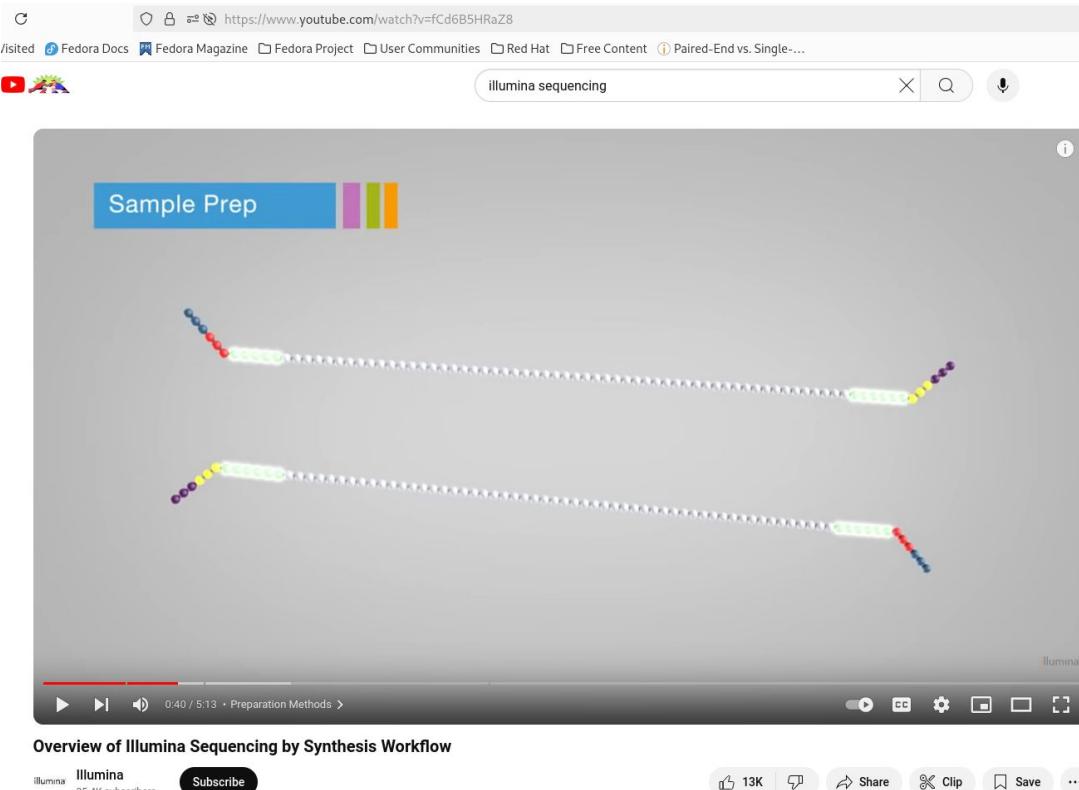
Save

...

# PCA – метод головних компонент

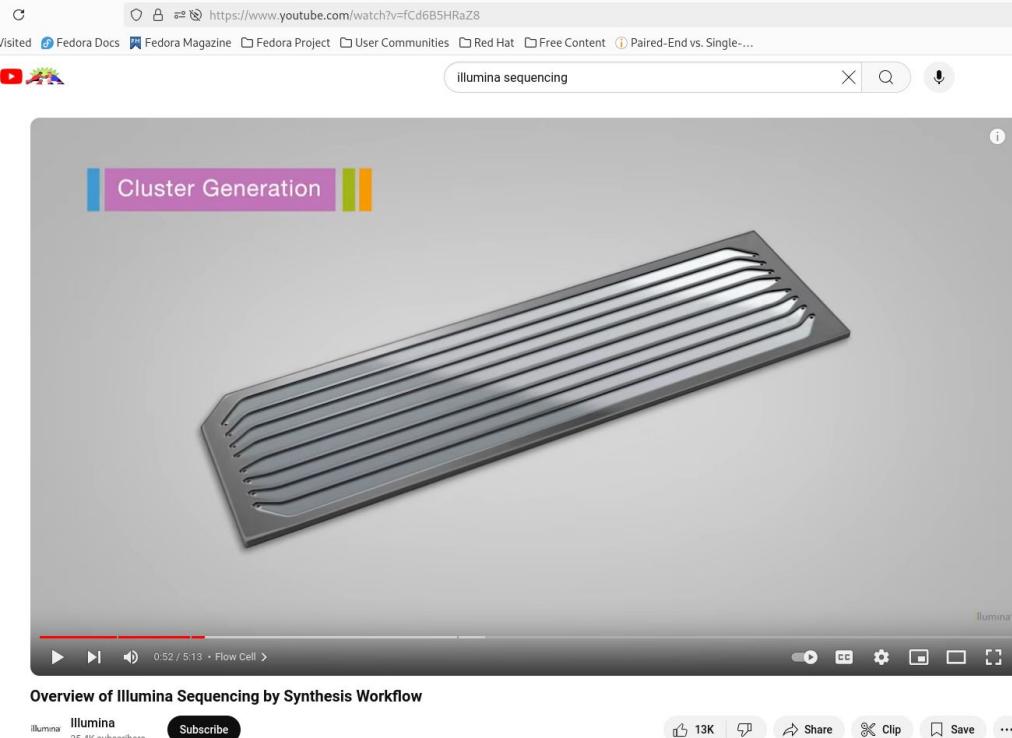
- [https://uk.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4\\_%D0%B3%D0%BE%D0%BB%D0%BE%D0%B2%D0%BD%D0%B8%D1%85\\_%D0%BA%D0%BE%D0%BC%D0%BF%D0%BE%D0%BD%D0%B5%D0%BD%D1%82](https://uk.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%B3%D0%BE%D0%BB%D0%BE%D0%B2%D0%BD%D0%B8%D1%85_%D0%BA%D0%BE%D0%BC%D0%BF%D0%BE%D0%BD%D0%B5%D0%BD%D1%82)
- PCA by Josh Starmer: <https://www.youtube.com/watch?v=FqakZw6K1QQ>
- PCA із прикладами в R із книги “Modern Statistics for Modern biology” by Susan Holmes and Wolfgang Huber  
<https://www.huber.embl.de/msmb/07-chap.html>

# Секвенування Next Generation Sequencing- Illumina



<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

# Секвенування Next Generation Sequencing- Illumina



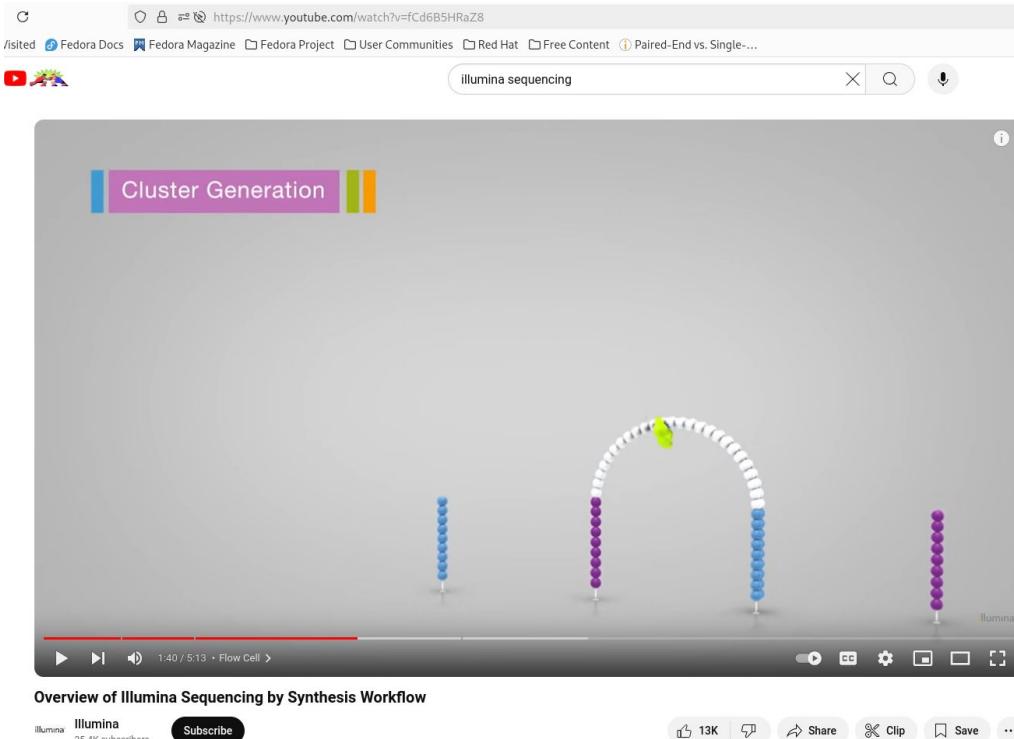
<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

# Секвенування Next Generation Sequencing - Illumina



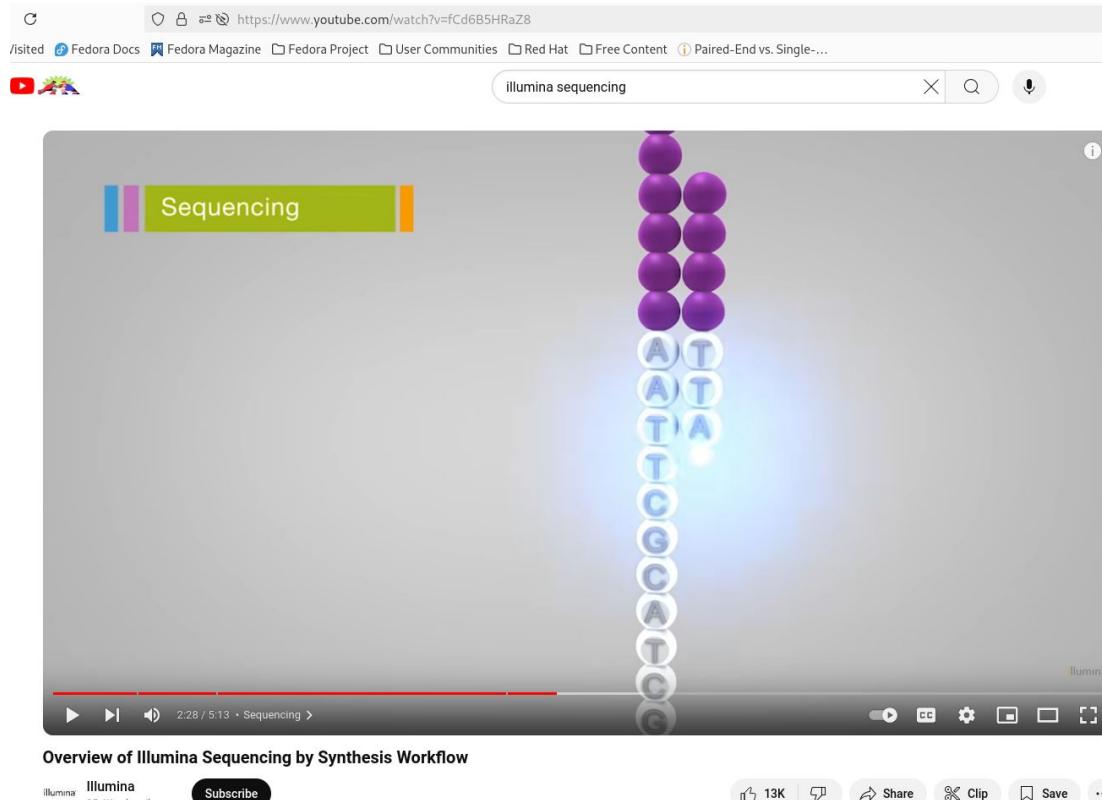
<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

# Секвенування Next Generation Sequencing- Illumina



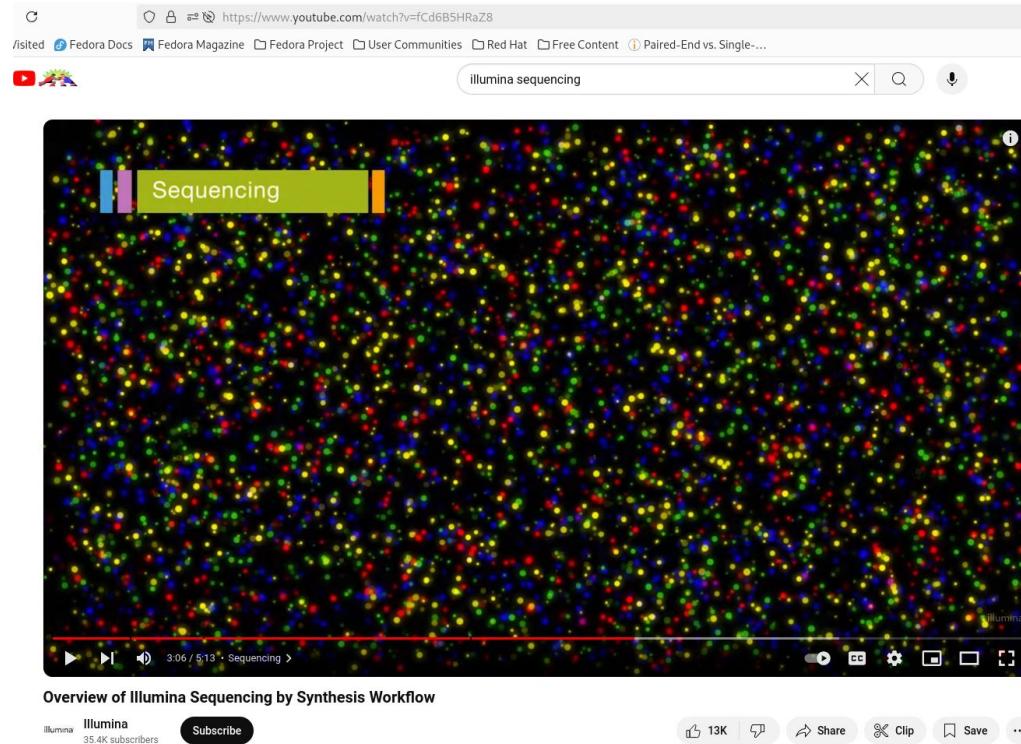
<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

# Секвенування Next Generation Sequencing - Illumina



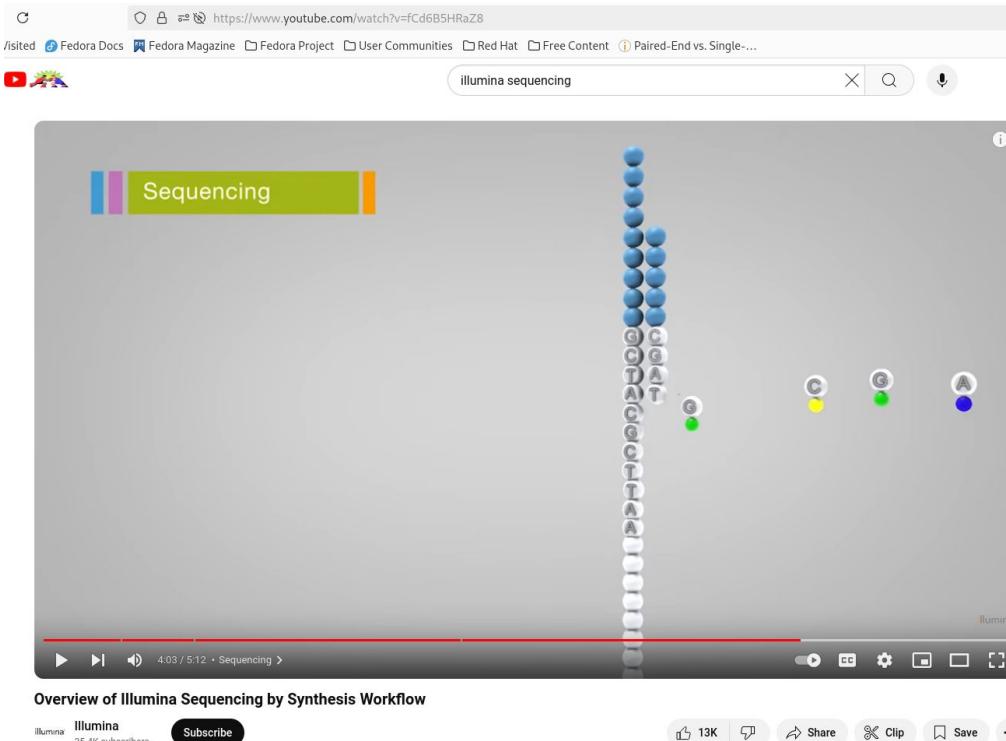
<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

# Секвенування Next Generation Sequencing- Illumina



<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

# Секвенування Next Generation Sequencing - Illumina



<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

# Секвенування Next Generation Sequencing- Illumina

The screenshot shows a web browser window with the URL <https://www.illumina.com/science/technology/next-generation-sequencing/mate-pair-sequencing.html>. The page title is "De Novo Assembly with Mate Pairs". The diagram illustrates the process: "Short-Insert Paired End Reads" (represented by orange and blue horizontal bars) and "Long-Insert Paired End Reads (Mate Pair)" (represented by green and purple horizontal bars). These are used to create a "De Novo Assembly" (represented by a network of connected horizontal bars of various colors). A text box explains: "Using a combination of short and long insert sizes with paired-end sequencing results in maximal coverage of the genome for de novo assembly. Because larger inserts can pair reads across greater distances, they provide a better ability to read through highly repetitive sequences and regions where large structural rearrangements have occurred. Shorter inserts sequenced at higher depths can fill in gaps missed by larger inserts sequenced at lower depths. Thus a diverse library of short and long inserts results in better de novo assembly, leading to fewer gaps, larger contigs, and greater accuracy of the final consensus sequence."

- [https://en.wikipedia.org/wiki/Illumina\\_dye\\_sequencing](https://en.wikipedia.org/wiki/Illumina_dye_sequencing)
- <https://www.youtube.com/watch?v=fCd6B5HRaZ8>

# P-value - р-значення - як інтерпретувати

- Тестуємо: зразки клітин печінки під впливом токсичної речовини (експеримент) і контроль
- Мета: знайти гени, експресія яких підвищена у експерименті, або знижена у експерименті (порівняно з контролем)
- $H_0$  (нульова гіпотеза/припущення) - експресія гена X у експерименті і контролі не відрізняються
- $H_1$  (тестова гіпотеза) - експресія гена X у експерименті відрізняється від контролю
- [https://uk.wikipedia.org/wiki/P-значення#/media/Файл:P-value\\_in\\_statistical\\_significance\\_testing\\_uk.svg](https://uk.wikipedia.org/wiki/P-значення#/media/Файл:P-value_in_statistical_significance_testing_uk.svg)
- $P\text{-value} = 1$  для гена A - “експресія відрізняється, але це 100% не точно”
- $P\text{-value} = 0.01$  для гена B
- $P\text{-value} = 1e-6$  для гена C

# Більше про р-значення – як рахувати

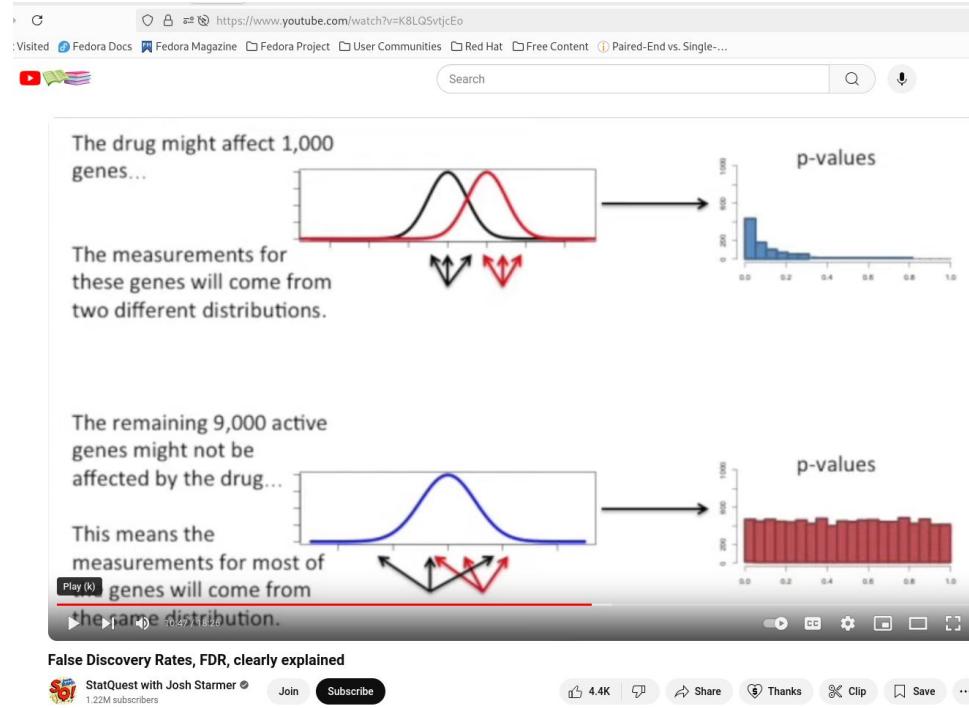
- Р-значення нічого не каже про амплітуду сигналу (Fold Change = FC) (наскільки відрізняється експресія)
- Дивіться відео
  - <https://www.youtube.com/watch?v=vemZtEM63GY>
  - <https://www.youtube.com/watch?v=JQc3yx0-Q9E>
- [https://github.com/GenomicsUA/2024-daad-rnaseq-course/blob/main/02\\_classes/04\\_rnaseq\\_intro\\_part2/05.pvalues.Rmd](https://github.com/GenomicsUA/2024-daad-rnaseq-course/blob/main/02_classes/04_rnaseq_intro_part2/05.pvalues.Rmd)

# Multiple testing correction – проблема багаторазового тестування

		Прогноз	
		Дощ	Сонячно
Реальність	Дощ	True Positive	False negative
	Сонячно	False Positive	True Negative

- ⌚ <https://www.youtube.com/watch?v=K8LQSvtjcEo>
- ⌚ Рівномірний розподіл

# Корекція багаторазового тестування



Рівномірний розподіл

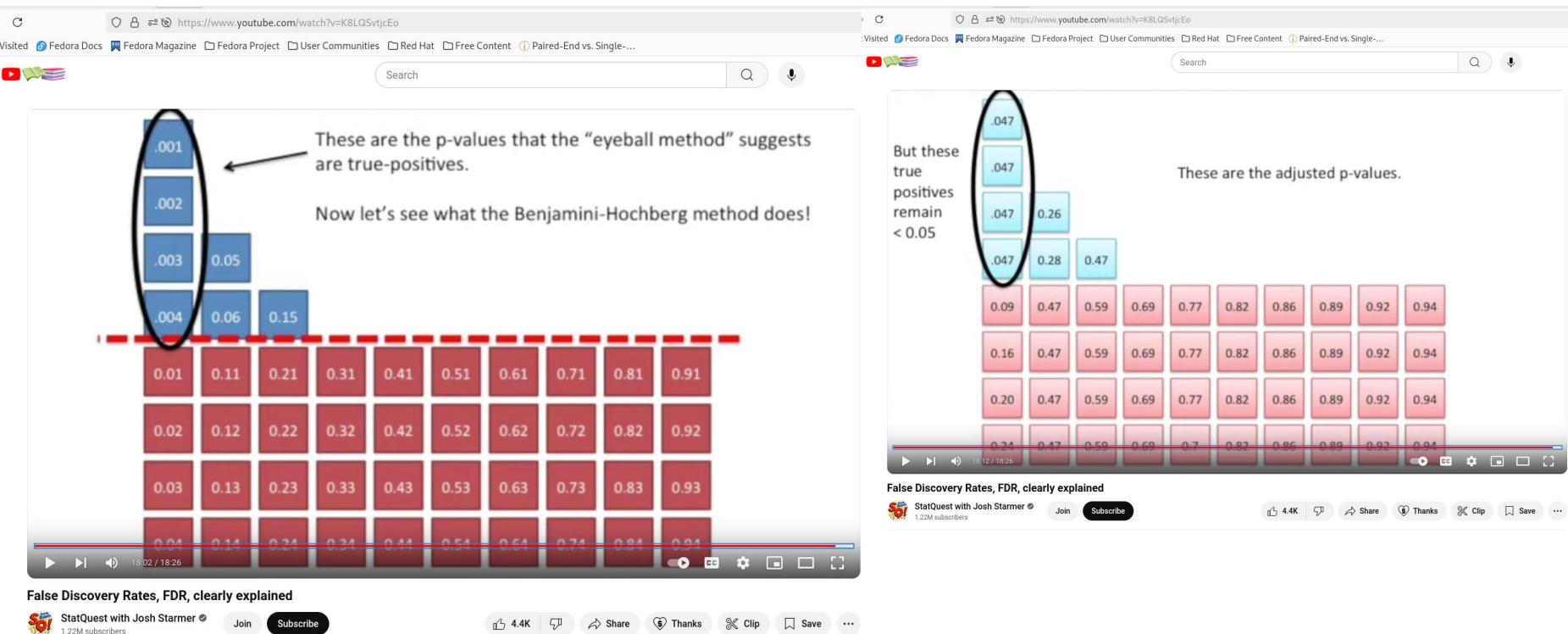


<https://www.youtube.com/watch?v=K8LQSvtjcEo>

# Корекція багаторазового тестування



# FDR adjusted p-values (метод Бенджаміні-Гогберга)



False Discovery Rates, FDR, clearly explained

StatQuest with Josh Starmer

Join

Subscribe

4.4K

Share

Thanks

Clip

Save