

# SCEXecute: Analize stratificate pe cod de bare celulare pentru date scRNA-seq

Analiză și prezentare articol științific

Mara Spataru   Alexandra Toma   Rareș Tudur

Universitatea București  
Inițiere în cercetare și bioinformatică

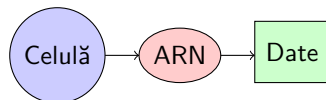
May 12, 2025

# Cuprins

- 1 Introducere în scRNA-seq
- 2 Coduri de bare celulare
- 3 Necesitatea analizelor stratificate pe cod de bare
- 4 Limitările instrumentelor tradiționale

# Ce este scRNA-seq?

- **Single-cell RNA sequencing** = secvențierea ARN la nivel de celulă unică
- Tehnologie care permite măsurarea expresiei genelor în fiecare celulă individual
- Revoluționează înțelegerea diferențelor între celule (nu mai lucrăm cu medii)
- Permite identificarea tipurilor rare de celule (care altfel ar fi "invizibile")



# RNA-seq tradițional vs. scRNA-seq

## RNA-seq tradițional:

- Analizează ARN din **grupuri** de celule
- Oferă o *medie* a activității genelor
- Pierde detaliile despre diferențele între celule
- Nu detectează evenimente care apar rar

## scRNA-seq:

- Analizează ARN din celule **una câte una**
- Arată diferențele dintre celule individual
- Descoperă tipuri rare de celule
- Urmărește evoluția diferitelor tipuri de celule
- Necesită algoritmi mai complecși

Comparație	RNA-seq	scRNA-seq
<b>Ce măsoară</b>	Grupe de celule	Celule individuale
<b>Ce observăm</b>	Doar media	Diferențele între celule
<b>Complexitate</b>	Mai simplă	Mai avansată

# Provocări în scRNA-seq

- **Tehnice:**

- Cantitate foarte mică de ARN într-o singură celulă
- Proces ineficient de izolare a ARN-ului (doar 10-40% este identificat)
- Multe gene nu sunt detectate deloc în unele celule (deși sunt prezente)
- **Zgomot tehnic** = erori și variații introduse de aparatură și proceduri

- **Computaționale:**

- Volume mari de date (sute de GB)
- Necesitatea identificării celulelor individuale din amestec
- Algoritmi specializați pentru **date rare/împrăștiate** (multe valori zero)
- Metode complexe pentru standardizare și comparare între celule

# Platforme principale scRNA-seq

Platformă	Tehnologie	Caracteristici
10x Genomics	Bazată pe picături	Volum mare de celule
Smart-seq2	Bazată pe plăci	Acoperire completă
Drop-seq	Bazată pe picături	Cost redus
MARS-seq	Bazată pe plăci	Automatizat

- Platforme bazate pe **picături** (droplet-based):
  - Procesează mii-zeci de mii de celule simultan
  - Necesită **coduri de bare** celulare pentru identificare
  - (Detalii oral: celulele sunt încapsulate în picături minuscule de apă în ulei)
- Platforme bazate pe **plăci** (plate-based):
  - Procesează sute de celule
  - Control mai bun, acoperire genomică mai bună
  - (Detalii oral: celulele sunt plasate în godeuri individuale pe o placă cu 96/384 poziții)

# Tehnologii scRNA-seq

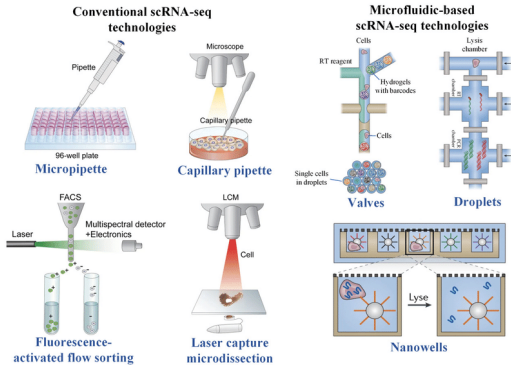
## Convenționale:

- Micropipetă (plăci 96/384)
- Capilar pentru celule unice
- FACS (sortare celule activate)
- Microdisecție cu laser

## Microfluidice:

- Droplets (picături cu celule)
- Valve microfluidice
- Nanogodeuri
- Debit mai mare de celule

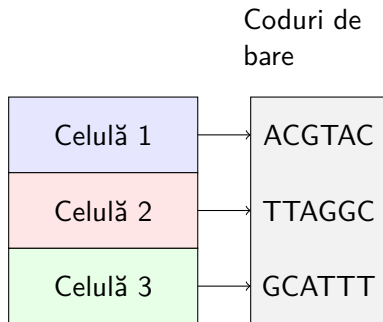
### Single-Cell RNA Sequencing (ScRNA-seq)



**Avantaj:** Combinare cu analiza proteinelor sau secvențierea genomului

# Ce sunt codurile de bare celulare?

- Secvențe scurte de nucleotide (10-20 baze/litere ADN)
- Atașate la ARN-ul din fiecare celulă individuală
- Acționează ca "etichete" pentru identificarea celulei de origine
- Esențiale pentru analiza a mii de celule simultan

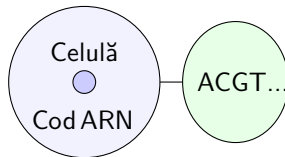




# Cum funcționează codurile de bare?

## 1 În tehnologiile bazate pe picături:

- Fiecare celulă este izolată într-o picătură minusculă
- Picătura conține un cod de bare specific
- ARN-ul extras din celulă primește acest cod de bare



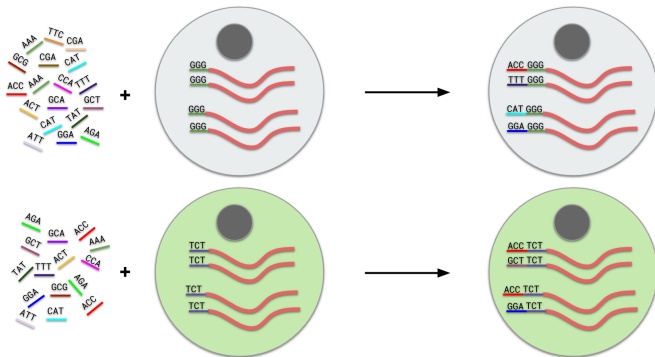
## 2 În tehnologiile bazate pe plăci:

- Celulele sunt izolate în compartimente separate
- Fiecare compartiment primește un cod de bare unic

## Avantaje:

- Permite analizarea a mii de celule într-un singur experiment
- Reduce semnificativ costurile per celulă
- Permite separarea datelor pe celule după secvențiere

# Mecanismul codurilor de bare în droplet-based scRNA-seq



Sursa: Galaxy Project (2022)

- Stânga: secvențe de coduri de bare | Mijloc: ARN din celule cu cod celular (GGG/TCT)
- Dreapta: ARN cu cod celular + UMI unic | Sistem dublu elimină duplicatele PCR

# Tipuri de coduri de bare în scRNA-seq

- **Cod de bare celular (Cell Barcode - CB):**

- Identifică din ce celulă provine secvența ARN
- Lungime: 10-20 litere/baze (A, C, G, T)
- În fișierele de aliniere, este stocat ca informație specială (tag)

- **Identificator Molecular Unic (UMI):**

- Identifică exact ce moleculă ARN a fost secvențiată
- Previne numărarea multiplă a aceluiași molecule (duplicate PCR)
- Lungime: 8-12 litere/baze

- **Index de probă:**

- Identifică din ce pacient/probă provine secvența
- Permite combinarea mai multor probe în același experiment

# Provocări în utilizarea codurilor de bare

## Provocări tehnice:

- Erori de citire în codurile de bare
- Coduri de bare identice pentru celule diferite
- Gene care nu sunt detectate deși există în celulă
- Eficiență variabilă la atașarea codurilor de bare

## Soluții:

- Algoritmi de corectare a erorilor de secvențiere
- Liste predefinite de coduri valide
- Filtrarea codurilor rare (probabil eronate)
- Stocarea informațiilor în câmpuri speciale în fișiere BAM

## Observație importantă

Datele scRNA-seq conțin toate celulele amestecate. Trebuie să separăm citirile (reads) pentru fiecare celulă folosind codurile de bare pentru a face analize la nivel de celulă.

# De ce sunt necesare analizele stratificate?

- **Expresia genică** - cel mai studiat aspect, dar nu singurul.
- Alte caracteristici la nivel de celulă sunt importante pentru:
  - Înțelegerea diferențelor dintre celule
  - Identificarea grupurilor rare de celule
  - Descoperirea modificărilor genetice rare
  - Studiul evoluției diferitelor grupe de celule
- **Stratificarea** = separarea citirilor (reads) pe celule individuale
- Permite aplicarea metodelor de analiză tradiționale la nivel de celulă

## Provocare computațională

Procesarea separată a datelor pentru sute/mii de celule necesită soluții eficiente.

# Informații adiționale din scRNA-seq (1/2)

## **Varianțe genetice (SNV-uri):**

- Modificări la nivelul unei singure baze în ADN/ARN
- Variații genetice specifice unei celule
- Mutații apărute după formarea embrionului (somatice)
- Modificări apărute după transcrierea ARN

## **Expresie diferențiată a alelelor (ASE):**

- Folosirea preferențială a unei versiuni a genei vs cealaltă
- Mecanism de control genetic
- Proces natural în inactivarea cromozomului X la femei
- Modificări în controlul expresiei genelor

# Informații adiționale din scRNA-seq (2/2)

## Splicing alternativ:

- Diferite moduri de asamblare a ARN-ului mesager
- Specific tipului de celulă
- Crește diversitatea proteinelor posibile
- Mecanism de reglare a funcțiilor celulare

## Editare ARN:

- Modificări chimice ale bazelor ARN
- Diferențe între ARN și ADN-ul original
- Variaza între tipuri de celule
- Crește diversitatea moleculelor ARN

# Cazuri de utilizare a analizelor stratificate

- **Cercetare în cancer:**

- Diferențe între celulele din aceeași tumoră
- Urmărirea evoluției grupurilor de celule
- Înțelegerea rezistenței la tratament

- **Imunologie:**

- Răspunsuri specifice ale celulelor imune
- Diversitatea receptorilor celulelor T și B

- **Neurobiologie:**

- Identificarea tipurilor de neuroni
- Modificări specifice în ARN-ul neuronal

- **Studiul dezvoltării organismelor:**

- Cum se specializează celulele din embrion în diferite țesuturi
- Cum "decid" celulele ce funcții să îndeplinească în organism



# Instrumente principale pentru scRNA-seq (1/2)

## STARsolo:

- Extensie a programului de aliniere STAR
- Combină alinierea cu identificarea celulelor
- Numără eficient genele exprimate
- Optimizat pentru date 10x Genomics

## CellRanger:

- Program dezvoltat de compania 10x Genomics
- Set complet de instrumente pentru scRNA-seq
- Combină aliniere + numărare + grupare
- Special creat pentru datele 10x

# Instrumente principale pentru scRNA-seq (2/2)

## UMI-tools:

- Procesează identificatorii moleculari unici
- Elimină duplicatele și numără corect
- Flexibil pentru diverse platforme

## Seurat/Scanpy:

- Instrumente pentru analiză și vizualizare
- Grupează celulele în tipuri similare
- Permite combinarea mai multor probe

# Limitări ale instrumentelor existente

- **Concentrare doar pe expresia genică:**
  - Optimizează pentru numărarea genelor
  - Nu oferă informații despre variante genetice, splicing, etc.
- **Lipsa suportului pentru analize personalizate:**
  - Nu putem folosi direct instrumente create pentru secvențierea clasică
  - Pentru fiecare tip nou de analiză, se dezvoltă programe speciale
- **Probleme de scalabilitate:**
  - Analiza separată a mii de celule este ineficientă
  - Consumă prea multă memorie pentru multe celule
- **Format de date neadecvat:**
  - Datele combinate (agregate) trebuie procesate suplimentar
  - Incompatibilitate cu programele existente

# Bariere în analiza avansată scRNA-seq

## Provocări tehnice:

- Separarea manuală a fișierelor BAM consumă timp
- Utilizarea inefficientă a resurselor de calcul
- Complexitate ridicată a fluxurilor de analiză
- Lipsa unei abordări standardizate

Focus limitat
Complexitate de calcul
Date combinate
Unelte inadecvate

## Nevoia unei soluții

Este nevoie de o metodă care să permită folosirea programelor existente pentru analiza genetică la nivel de celulă, fără a consuma prea multe resurse.

# Problema tehnică la care răspunde SCExecute

- Programele actuale pentru scRNA-seq generează fișiere BAM cu toate celulele
- Programele clasice de analiză genetică (GATK, Strelka2) nu înțeleg codurile de bare
- Este nevoie de:
  - Extragerea secvențelor pentru fiecare celulă
  - Crearea de fișiere BAM separate per celulă (scBAM)
  - Aplicarea programelor de analiză pe fiecare scBAM
  - Combinarea rezultatelor pentru interpretare
- SCExecute este creat exact pentru a rezolva această problemă.

**SCExecute** permite utilizarea instrumentelor existente pentru analiza datelor la nivel de celulă.

# Concluzii

- scRNA-seq oferă rezoluție celulară pentru studierea expresiei genelor
- Codurile de bare celulare sunt esențiale pentru identificarea originii celulare a secvențelor
- Analizele stratificate pe cod de bare permit studiul:
  - Variantelor genetice exprimate (SNV)
  - Expresiei diferențiate a alelelor
  - Splicing-ului alternativ
  - Editării ARN
- Programele existente pentru scRNA-seq sunt limitate la analiza expresiei genelor
- SCExecute este soluția pentru analize personalizate la nivel de celulă