COSMOTECH

Single-cell RTq-PCR: Raw data treatment

Arnaud Bonnaffoux

Single-Cell day

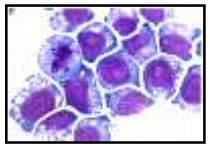
ENS DE LYON

LBMC

Inria
INVENTEURS DU MONDE NUMÉRIQUE
Équipe DRACULA

anrt
association nationale
recherche technologie

INTRODUCTION



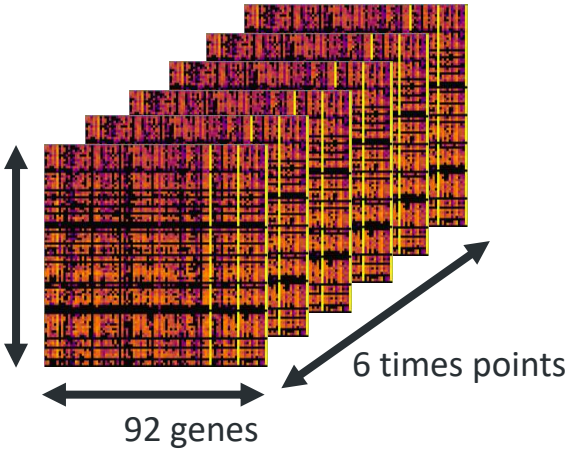
Your cells

Technical « noise »

BioMark Fluidigm
Single Cell RTqPCR data

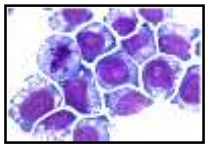
Raw data treatment

Single-cell
RNA counts matrices

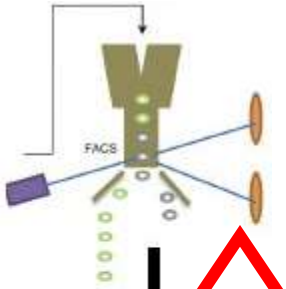


(Richard et al. (2016) PLoS Biology)

SOURCE OF TECHNICAL NOISE



Your cells



DROP OUT

Spike 1
Spike 2

+ Lysis buffer

Retro Transcription

HIGH NOISE

Pre-amplification (20 cycles)

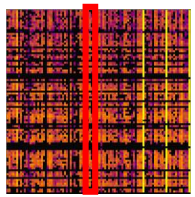
qPCR

LOW NOISE

Raw data treatment/checks steps

1. Deletion of abnormal amplifications (check qPCR)
2. « Black cells » deletion (check drop-outs)
3. Deletion of low-detected gene (check RT)
4. Spike normalisation (RT bias correction)
5. Conversion of Ct in molecule counts
6. Correction factor (dilution bias correction)
7. Gene triplicate (check qPCR)

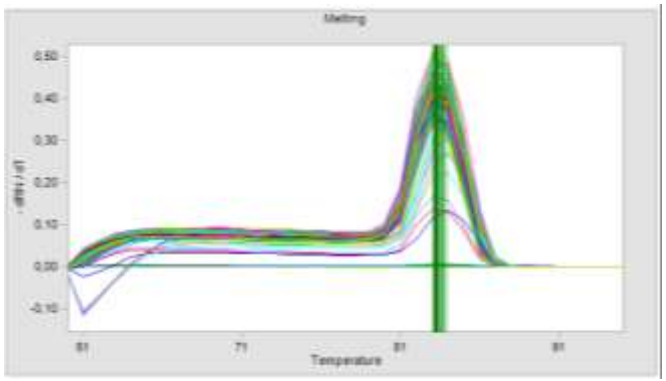
1- Deletion of abnormal amplified cell/gene



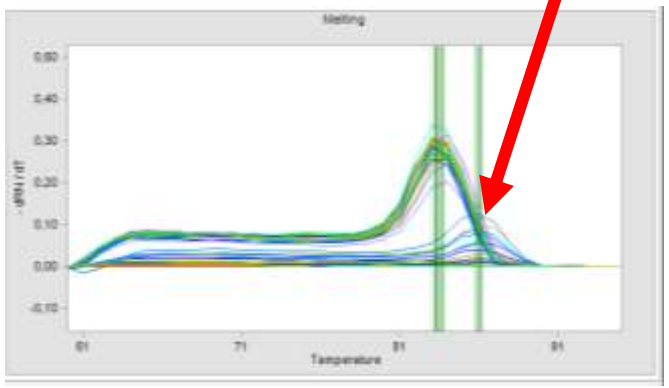
Single Cell

RT-qPCR raw data

OK



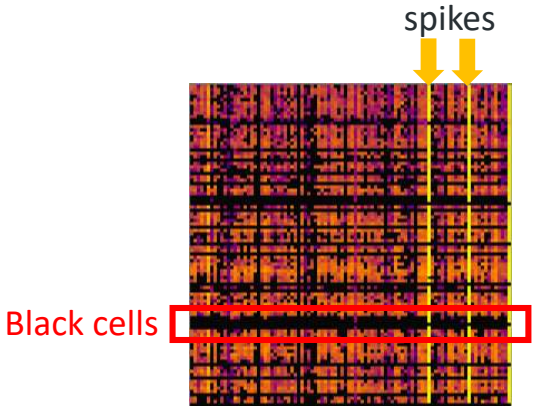
deleted

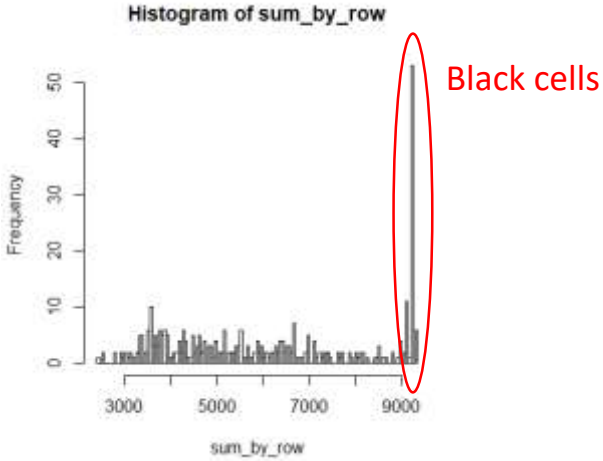


- Export Ct data in .csv file
- Observe « manually » melt curve of one gene for all cells using Fluidigm software
- If a pic is shifted (delta > 0.6°) compared to main pic, set Ct value for couple cell/gene to UD (UnDefined)

2- « Black » cells deletion

- « Black » cells = Spikes with low Ct, all other genes have imporant Ct (Ct > 30 ≠ UnDefined)
- Even if genes expression is bursty (important probability of null value), the probability to have ALL genes to 0 is very low
- Proportion of black cells can be important :
  - AR11 : 27%
  - AR78 : 54%
  - AR85 : 32%
- Check absence of « black » cells using sum of Ct per cell

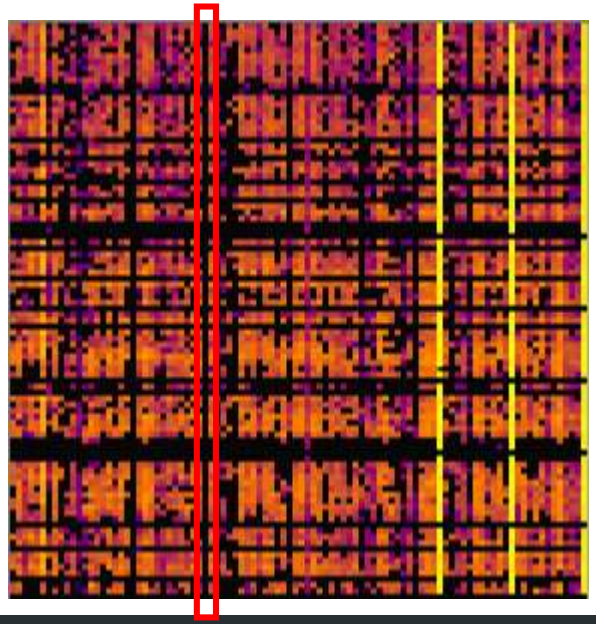




AR11

3- Deletion of low-detected gene

- For genes with important UnDefined values (≠ 0) in cells (> 80%), sample size is too low to be considered
- Only few genes are concerned
  - AR78 : 3 / 93
  - AR85 : 2 / 93



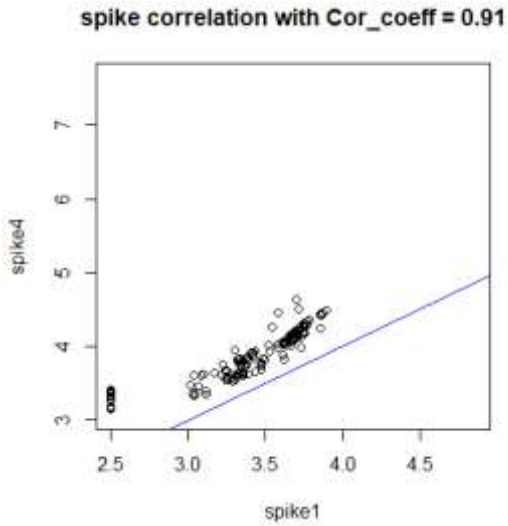
4- SPIKE : check RT + normalisation

- At least one spike shall be detected and Ct < 30, else cell is deleted
- If spikes correlation is OK (R>0,7) normalise Ct values of a cell with its spike values
- Spike correlation is « stochastic » (???)

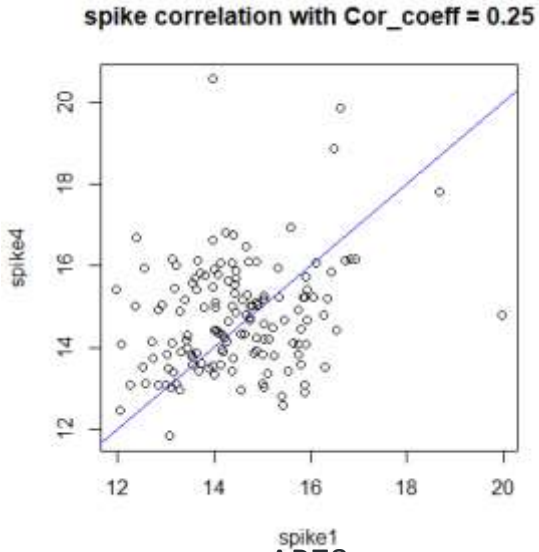
Raw Ct — All cells mean spikes Ct

$$Cq_{i,j} = \widehat{Cq}_{i,j} - \left(\overline{Cq}_i - \overline{Cq}_0\right)$$

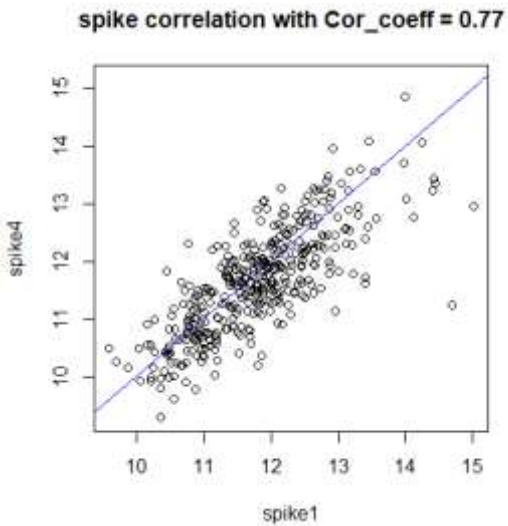
Ct normalized — Spike Ct



AR11



AR78



AR85

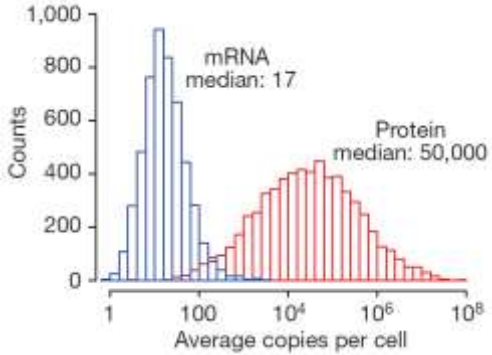
5- Conversion of Ct in molecule count

$$\mathrm{m}_{i,j} = 96 \times 45 \times 2^{30-22-\mathrm{Cq}_{i,j}}$$

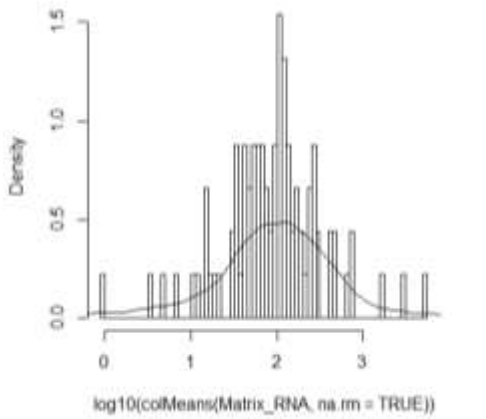
- 30 = detection threshold
- 22 = cycles of pre-amplification
- 96 = dilution in 96 wells
- 45 = sampling factor for qPCR measurement

6- Correction factor: Total distribution centering

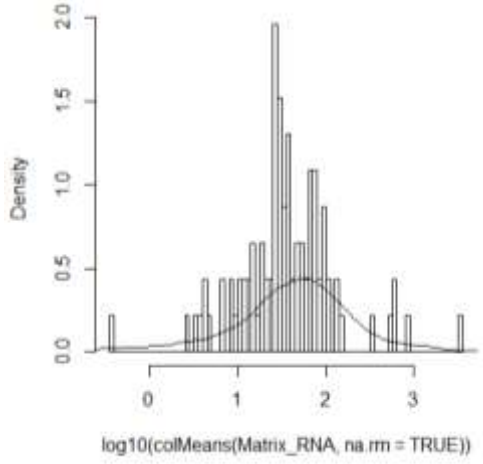
- To correct possible bias due to imprecision of dilution factor or detection threshold compare your total gene expression distribution (in all cells) with litterature distribution



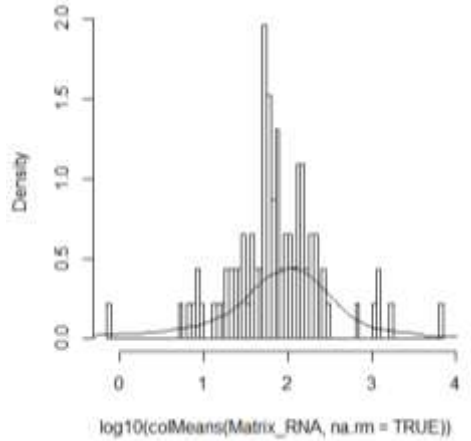
Schwanhausser et al., 2013, Nature



AR78 (*1)



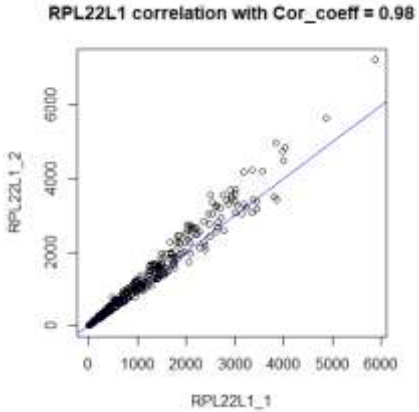
AR85 (*1)

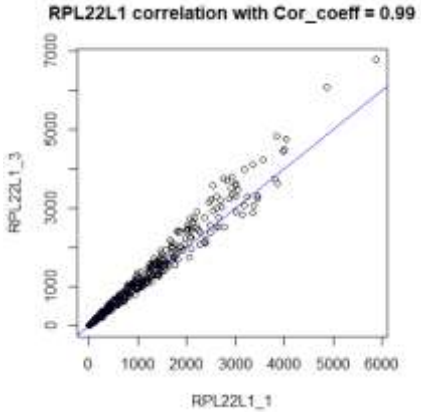


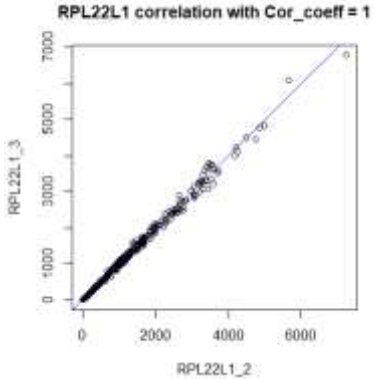
AR85 (*2)

7-GENE TRIPLICATE : check qPCR

- Use 3 identical primers for one gene (RPL22L1)
- Usually qPCR replication is quite perfect







Raw data treatment/checks steps

1. Deletion of abnormal amplifications (check qPCR)
2. « Black cells » deletion (check drop-outs)
3. Deletion of low-detected gene (check RT)
4. Spike normalisation (RT bias correction)
5. Conversion of Ct in molecule counts
6. Correction factor (dilution bias correction)
7. Gene triplicate (check qPCR)