

Patel et al. Science 2014

Part 1

1 RNA-seq data

2 Bowtie (1.4.1)

-n 0 -e 999999999 -l 25 -I 1 -X 2000 -a -m 15 -S

3 RSEM (1.2.3)

estimate-rspd

4 TPM

5 $\log_2(\text{TPM}+1)$

6 filter genes

with average $\log_2(\text{TPM}+1) > 4.5$ across all cells

or average $\log_2(\text{TPM}+1) > 6$ in at least one tumor

7 filter cells

At least 2000 genes expressed

8 relative expression levels

$= \log_2(\text{TPM}+1) - \text{average } \log_2(\text{TPM}+1) \text{ of all cells}$

9 sorted genes

10 $\text{CNV}_k(i) = \text{SUM} (j \text{ from } i-50 \text{ to } i+50) E_k(oj) / 101$

11 center CNV vector at zero for each cell

Part 2

12 normal sample TPM (from count <http://www.gtexportal.org/>)

13 $\log_2(\text{TPM}+1) - \text{average } \log_2(\text{TPM}+1) \text{ of all tumor single cells}$

14 reference CNV vector (normal) CNVcont

Part 3

15 cluster (average-linkage hierarchical clustering)

16 average CNV profile of the 'normal' cluster one as a reference

17 subtracting reference from the CNV profiles

```
tpm ← function( counts, lengths) {  
  rate ← counts / lengths  
  return( rate / sum(rate) * 1e6 )  
}
```

aggregated expression

$= \log_2(\text{average TPM of all single cell from the same tumor} + 1)$

Tirosh et al. Science 2016

- 1 Limit relative expression values to $[-3, 3]$
- 2 $\text{CNV} - \text{BaseMax}$, if $\text{CNV} > \text{BaseMax} + 0.2$
 $\text{CNV} - \text{BaseMin}$, if $\text{CNV} < \text{BaseMin} - 0.2$
0, if $\text{BaseMin} - 0.2 < \text{CNV} < \text{BaseMin} + 0.2$

Darmanis et al. Cell Report 2017

- 1 The CNV vector is a moving average of gene expression using a window of $0.1 \times n$ genes per chromosome
- 2 The resulting CNV vectors of each cell were centered by subtraction of their mean prior to any downstream analysis