More Discussion about the “3:1” or “1:3” Ratio

Bokai Wang

12/14/2015

Index of this Presentation

- Introduction of the "3:1" or "1:3" Ratio
- Investigation of the "3:1" or "1:3" Ratio (OLSE)
- Investigation of the "3:1" or "1:3" Ratio (WLSE)
- Conclusion of the OLSE and the WLSE
- Any questions or suggestions

Intro. of the “3:1” or “1:3” Ratio

- The well-characterized reference RNA samples:
- A: Universal Human Reference RNA
- B: Human Brain Reference RNA
- A and B both from the MAQC consortium, adding spike-ins of synthetic RNA from the External RNA Control Consortium (ERCC).
- C: mixed A and B in ratio 3:1
- D: mixed A and B in ratio 1:3

Intro. of the "3:1" or "1:3" Ratio

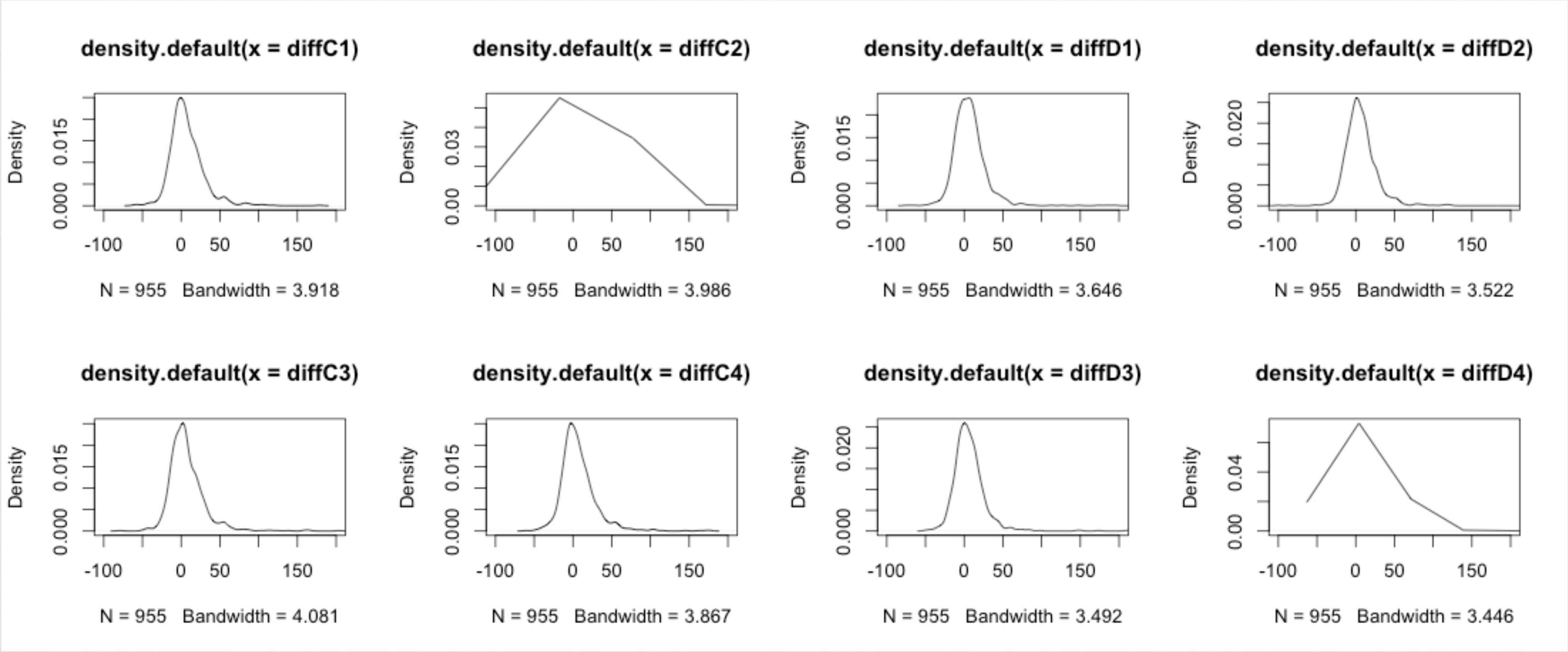
- RNA samples A, B, C, and D were analyzed by using the TaqMan RT-PCR technology in the SEQC project.

- But I found the expression intensity of samples C and D may not be a linear combination ("3:1" or "1:3") of A and B:

```
> (3*taqman$A.A1_value[1]+taqman$B.B1_value[1])/4; taqman$C.C1_value[1]
[1] 0.007067388
[1] 0.00722122
> (taqman$A.A1_value[1]+3*taqman$B.B1_value[1])/4; taqman$D.D1_value[1]
[1] 0.003977883
[1] 0.00308095
```

Intro. of the "3:1" or "1:3" Ratio

- Investigate the difference in percent of the 4 replicates:



Intro. of the "3:1" or "1:3" Ratio

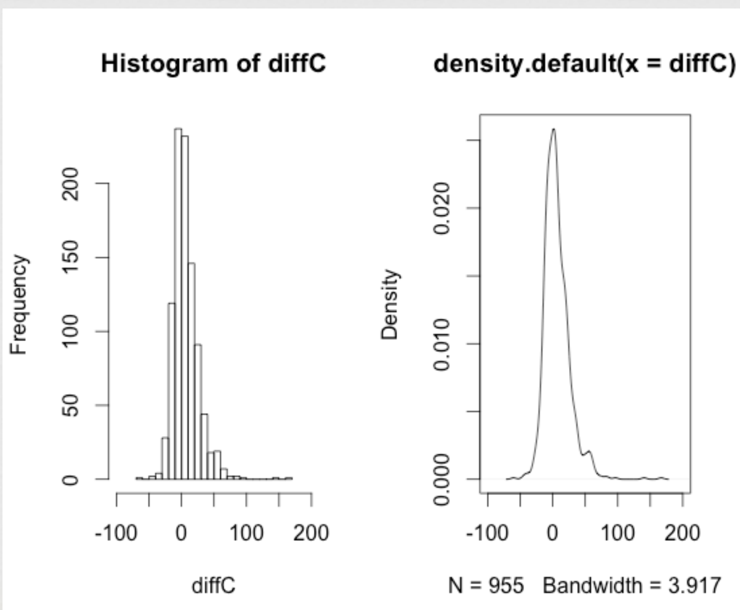
- Investigate the difference in percent of the 4 replicates:

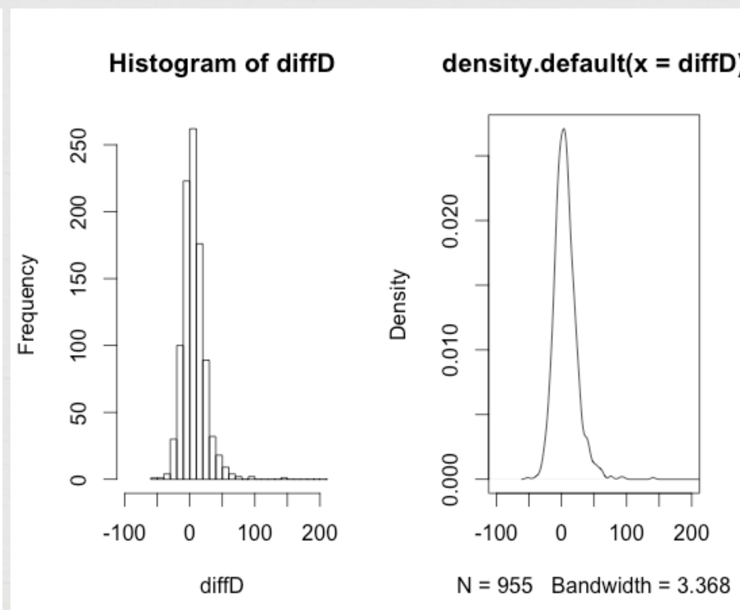
```
> ## show the sorted diffC2...
> diffC2[order(diffC2, decreasing = T)]
 [1] 4.819449e+04 1.801919e+02 1.515685e+02 9.938433e+01 9.665229e+01 9.567301e+01 9.154578e+01 8.975632e+01
 [9] 8.974486e+01 8.771303e+01 8.053520e+01 8.018463e+01 7.775564e+01 7.542093e+01 7.386648e+01 7.371656e+01
[17] 7.189206e+01 7.159055e+01 6.784994e+01 6.701318e+01 6.164792e+01 6.080071e+01 5.766504e+01 5.619394e+01
```

```
> ## show the sorted diffD4...
> diffD4[order(diffD4, decreasing = T)]
 [1] 3.432681e+04 6.097230e+02 3.548568e+02 1.671805e+02 1.662612e+02 1.509512e+02 1.284149e+02 1.095764e+02
 [9] 1.008479e+02 9.143596e+01 9.060098e+01 8.813691e+01 8.257313e+01 7.248578e+01 7.102093e+01 6.740137e+01
[17] 6.047998e+01 5.727314e+01 5.690205e+01 5.637384e+01 5.343375e+01 5.336175e+01 5.066168e+01 4.979269e+01
```

Intro. of the "3:1" or "1:3" Ratio

- Take the average of 4 samples A, B, C and D, and check to see if the issue above still exists, then investigate the difference in percent of the averages:





Intro. of the "3:1" or "1:3" Ratio

- Why take averages can solve the issue…?

```
> ## show the sorted diffC...
> diffC[order(diffC, decreasing = T)]
 [1] 166.83701011 140.56035895  96.42478253  84.46308209  80.81684517  74.08002249  71.03332474  66.87056334
 [9]  64.95494936  62.26121348  61.48919672  61.39311612  60.37565798  60.34416625  59.19792001  58.56319810
[17]  58.46625187  58.45027004  57.28222342  57.23029939  56.80219322  55.92598413  55.52087549  55.52087549
```

```
> ## show the sorted diffD...
> diffD[order(diffD, decreasing = T)]
 [1] 663.35096696 140.71589622  95.00632914  90.54767202  76.46643796  75.48043366  64.29595560  62.60191457
 [9]  60.94276289  60.41560119  59.98406718  58.92306254  56.09710844  55.21350200  54.06889357  53.83216078
[17]  53.39663712  52.22357351  50.99826860  49.44171391  49.43684286  48.76180152  47.80823513  47.50935031
```

Invest. of the "3:1" or "1:3" Ratio (OLSE)

- Instead of "3:1" (or "1:3"), assume the "true" ratio is "a:b", thus the expression of sample C = a/(a+b)*A + b/(a+b)*B (assuming "independence")
- For sample C, let's start with using simple linear regression to solve a by fixing b=1, then get a 95% CI about a
- Similarly, still for sample C, using simple linear regression to solve b by fixing a=3, then get a 95% CI about b

Invest. of the "3:1" or "1:3" Ratio (OLSE)

- Like before, but for sample D, let's start with using simple linear regression to solve a by fixing b=3, then get a 95% CI about a

- Similarly, but still for sample D, using simple linear regression to solve b by fixing a=1, then get a 95% CI about b

Invest. of the "3:1" or "1:3" Ratio (OLSE)

- The (adjusted) 95% CIs for a and b mentioned before are below
- For sample C:

```
> ## show the 95% CI for a in this case
> 4*confint(slfit1, level = 0.95)
               2.5 %   97.5 %
exp.A_value 3.041532 3.086699
```

```
> ## show the 95% CI for b in this case
> 4*confint(slfit2, level = 0.95)
               2.5 %    97.5 %
exp.B_value 0.847477 0.8962114
```

- For sample D:

```
> ## show the 95% CI for a in this case
> 4*confint(slfit3, level = 0.95)
               2.5 %   97.5 %
exp.A_value 1.005404 1.042664
```

```
> ## show the 95% CI for b in this case
> 4*confint(slfit4, level = 0.95)
               2.5 %   97.5 %
exp.B_value 2.805223 2.840509
```

Invest. of the "3:1" or "1:3" Ratio (OLSE)

- If doubt about the SLR method before, then try to use Multiple Linear Regression to solve a and b both, then get (adjusted) 95% CIs about them
- For sample C:

```
> ## construct the 95% CI for the case when a=3? and b=1?
> 4*confint(mlfit1, level = 0.95)[,2]/sum(confint(mlfit1, level = 0.95)[,2])
exp.A_value exp.B_value
  3.1518811   0.8481189
> 4*confint(mlfit1, level = 0.95)[,1]/sum(confint(mlfit1, level = 0.95)[,1])
exp.A_value exp.B_value
  3.1822451   0.8177549
```

- For sample D:

```
> ## construct the 95% CI for the case when a=1? and b=3?
> 4*confint(mlfit2, level = 0.95)[,2]/sum(confint(mlfit2, level = 0.95)[,2])
exp.A_value exp.B_value
   1.135584    2.864416
> 4*confint(mlfit2, level = 0.95)[,1]/sum(confint(mlfit2, level = 0.95)[,1])
exp.A_value exp.B_value
   1.122458    2.877542
```

Invest. of the "3:1" or "1:3" Ratio (WLSE)

- How to choose an appropriate weights?

- For sample C, want to minimize the percent^2

  $= ((a/(a+b)*A + b/(a+b)*B - C)/C)^2$

  $= 1/(C^2)*(a/(a+b)*A + b/(a+b)*B - C)^2$

  (This WLSE = The OLSE of min. percents)

Invest. of the "3:1" or "1:3" Ratio (WLSE)

- However, it might be more reasonable to use the weighted least squares
- For sample C:

```
> ## show the 95% CI for a in this case
> 4*confint(wslfit1, level = 0.95)
               2.5 %   97.5 %
exp.A_value 2.774606 2.874304
```

```
> ## show the 95% CI for b in this case
> 4*confint(wslfit2, level = 0.95)
                2.5 %    97.5 %
exp.B_value 0.7283639 0.7651916
```

- For sample D:

```
> ## show the 95% CI for a in this case
> 4*confint(wslfit3, level = 0.95)
                2.5 %    97.5 %
exp.A_value 0.7539548 0.8175869
```

```
> ## show the 95% CI for b in this case
> 4*confint(wslfit4, level = 0.95)
               2.5 %   97.5 %
exp.B_value 2.504119 2.629284
```

Invest. of the "3:1" or "1:3" Ratio (WLSE)

- Like before, then try to use Weighted Multiple Linear Regression to solve a and b both, then get (adjusted) 95% CIs about them
- For sample C:

```
> ## construct the 95% CI for the case when a=3? and b=1?
> 4*confint(wmlfit1, level = 0.95)[,2]/sum(confint(wmlfit1, level = 0.95)[,2])
exp.A_value exp.B_value
  3.1974997   0.8025003
> 4*confint(wmlfit1, level = 0.95)[,1]/sum(confint(wmlfit1, level = 0.95)[,1])
exp.A_value exp.B_value
  3.2136672   0.7863328
```

- For sample D:

```
> ## construct the 95% CI for the case when a=1? and b=3?
> 4*confint(wmlfit2, level = 0.95)[,2]/sum(confint(wmlfit2, level = 0.95)[,2])
exp.A_value exp.B_value
  0.9715463   3.0284537
> 4*confint(wmlfit2, level = 0.95)[,1]/sum(confint(wmlfit2, level = 0.95)[,1])
exp.A_value exp.B_value
  0.9500945   3.0499055
```

Conclusion of OLSE and WLSE

- C: Coverage
- N: Not Covered

- For OsLSE, for sample C & D
- a=3 (almost C) b=1 (N)
- a=1 (almost C) b=3 (N)

- For OmLSE, for sample C & D
- **a=3 & b=1 (N)**
- **a=1 & b=3 (N)**

Conclusion of OLSE and WLSE

- For WsLSE, for sample C & D
- a=3 (N) b=1 (N)
- a=1 (N) b=3 (N)

- For WmLSE, for sample C & D
- **a=3 & b=1 (N)**
- **a=1 & b=3 (almost C)**

Conclusion of OLSE and WLSE

- For OsLSE, for sample C
- a=3 (adjR^2=0.9867, AIC=-161.76, BIC=-152.04)
- b=1 (adjR^2=0.8377, AIC=-232.28, BIC=-222.56)

- For OsLSE, for sample D
- a=1 (adjR^2=0.9241, AIC=-529.34, BIC=-519.62)
- b=3 (adjR^2=0.9904, AIC=-848.97, BIC=-839.25)

- For OmLSE, for sample C & D
- a=3 & b=1 (adjR^2=**0.9913**, AIC=**-342.52**, BIC=**-327.94**)
- a=1 & b=3 (adjR^2=**0.9943**, AIC=**-976.20**, BIC=**-961.62**)

Conclusion of OLSE and WLSE

- For WsLSE, for sample C
- a=3 (adjR^2=0.9283, AIC=-4940.68, BIC=-4930.95)
- b=1 (adjR^2=0.8690, AIC=**-5435.75**, BIC=**-5426.03**)

- For WsLSE, for sample D
- a=1 (adjR^2=0.7109, AIC=-4604.51, BIC=-4594.79)
- b=3 (adjR^2=0.8715, AIC=-4612.94, BIC=-4603.22)

- For WmLSE, for sample C & D
- a=3 & b=1 (adjR^2=**0.9756**, AIC=-5433.89, BIC=-5419.30)
- a=1 & b=3 (adjR^2=**0.9382**, AIC=**-4687.01**, BIC=**-4672.43**)

Conclusion of OLSE and WLSE

- Thus, apply the Model Selection criteria in the last two pages, the conclusion is only when minimize the percent^2 of sample D, we might trust the ratio “1:3”.

- However, when minimize the expression^2 of sample C and D, and when minimize the percent^2 of sample C, we cannot trust “3:1” or “1:3”.

- Thus, what I am feeling is that ratio “3:1” for sample C cannot be trusted, but ratio “1:3” for sample D might be reasonable.

Any Questions or Suggestions