

Bioinformatik Übung 7 20.6.2022 Mariana Leisner

1. Quantil Normalisierung

	X			X _{sort}			X _{sort} rank(X _{sort})		rank(X)			X _{norm}		
	A	B	C	A	B	C			A	B	C	A	B	C
G ₁	22	2	3	17,5	0,5	0	6	i	vi	vi	v	9,6	9,6	8,16
G ₂	20	1	1	18	0,5	0	6,16	ii	v	iii	iii	8,16	7	7
G ₃	18	0,5	0	19	1	1	7	iii	ii	i	i	6,16	6	6
G ₄	19	1	5	19	1	1	7	iv	iii	iv	vi	7	7	9,6
G ₅	19	2,5	1	20	1,5	3	8,16	v	iv	vii	iv	7	10,83	7
G ₆	23	1,5	7	22	2	5	9,6	vi	vii	v	vii	10,83	8,16	10,83
G ₇	17,5	0,5	0	23	2,5	7	10,83	vii	i	ii	ii	6	6,16	6,16

iii. & iv. Rang. kriegen gemeinsam den 3,5. Rang

2. Student t-Test

$$T\text{-value} = \frac{\text{mean A} - \text{mean B}}{\frac{(n_A - 1) \cdot \text{var}_A^2 + (n_B - 1) \cdot \text{var}_B^2}{n_A + n_B - 2} \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

A | 42 39 38 60 41

mean A = 44

B | 38 42 56 64 68 69 62

mean B = 57

$n_A = 5$

$n_B = 7$

Variant(A): 2, 5, 6, 16, 3 $\xrightarrow{\text{quadrieren}}$ 4, 25, 36, 256, 9

Variant(B): 19, 15, 1, 7, 11, 12, 5 \rightarrow 361, 225, 1, 49, 121, 144, 25

$$S = \sqrt{\frac{4 + 25 + 36 + 256 + 9 + 361 + 225 + 1 + 49 + 121 + 144 + 25}{5 + 7 - 2}} = \sqrt{125,6} \approx 11,21$$

$$T\text{-value} = \frac{44 - 57}{11,21 \cdot \sqrt{\frac{1}{5} + \frac{1}{7}}} = -1,98$$

3. Bonferroni-Korrektur

Idee: wenn ich mehrere Tests auf den selben Daten mache, akkumuliert sich meine Fehlerwahrscheinlichkeit (wird größer) \rightarrow muss korrigiert werden
(Bei Vergleich von > 4 Gruppen: stattdessen lieber Türkei-test) für Bachelorarbeit

$$\frac{0,05}{100} = 5 \cdot 10^{-4}$$

DEGs mit p-Wert $\leftarrow \frac{\alpha}{\# \text{Gene}}$ sind signifikant werden akzeptiert

$$\Rightarrow p(\text{Gene}) \leftarrow 5 \cdot 10^{-4}$$

nicht signifikante / akzeptierte Gene: COX1, TFAM, ~~BRCA1, BRCA2, p53~~
0,04 10^{-3}

signifikante Gene: Her2, Ras, BRCA1, BRCA2, p53

Wahrscheinlich wurden gesunde mit Krebszellen verglichen
(Her2, Ras, BRCA1 & 2, p53 sind oft mutiert)

4. bei 0,01

1. 141 von 15952 Genen sind erhöht $\rightarrow 0,88\%$
($LFQ > 0$ (up))
2. Signifikanzschwelle größer \rightarrow mehr bunte Punkte
" kleiner \rightarrow weniger bunte Punkte

3. erhöhte Gene

res.order(res\$log2FoldChange, decreasing = TRUE)

[A] Sobie. 0080162600 : 5,45543

[2] Sobic, 006 G070564 : 5,26315

[3] Sobic.010G082600 : 4,73206

weggelassen

unterdrückte Gene

[1] Sabic, 0040101400 : - 7,14179

[2] Sobic. 009 G 182400 : -3,75887

[3] Sobic. 0016012200 : -3,68125

4. 1. kleinsten padj-Wert identifizieren: $4,75154 \cdot 10^{-16}$
(`order(res$padj)`)
als Signifikanz einsetzen \rightarrow kein Gen liegt mehr
drüber / drunter \rightarrow keine signifikanten Gene mehr