Differential gene expression analysis : Example of RNAseq

- Use statistics to compare 2 groups:

  For each gene i, is there a significant difference in mean expression between control and patients?

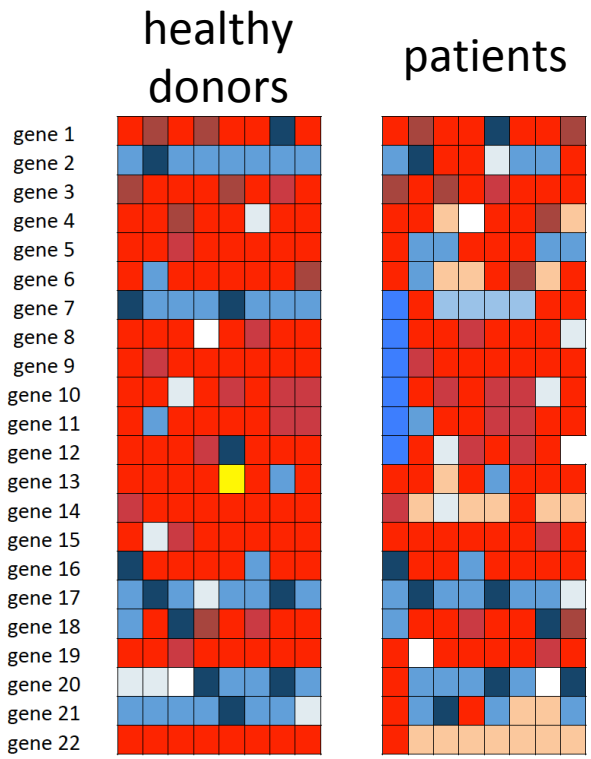
- T-test:

H0: Healthy donors and patients have similar gene I expression

H0i : $\pi i1 = \pi i2$

H1: Healthy donors and patients don't have a similar gene i expression

H1i : $\pi i1 \neq \pi i2$



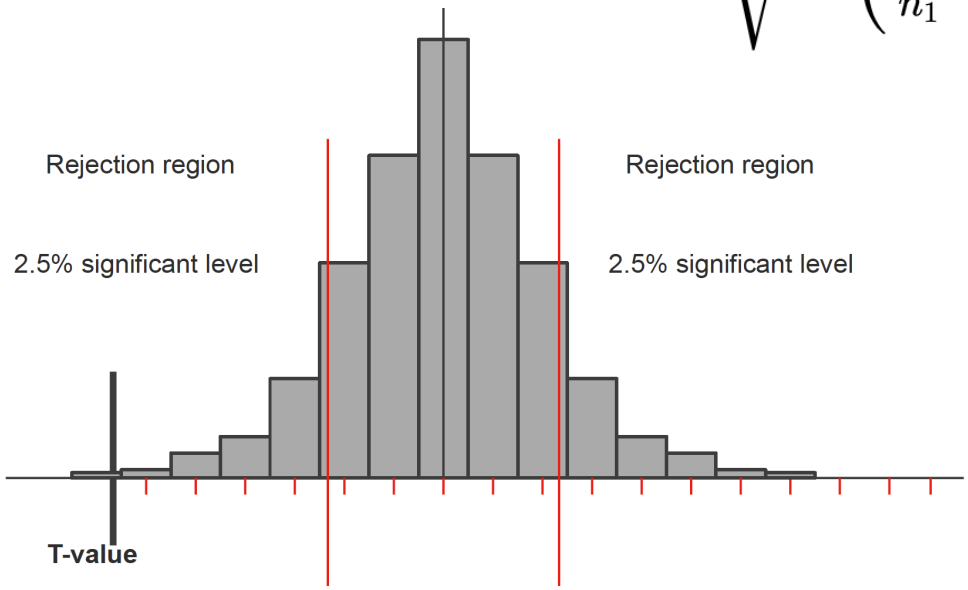
T-test in R

```
> t.test(grp1, grp2, paired = F)

        Welch Two Sample t-test

data:  grp1 and grp2
t = -6.3689, df = 8.9195, p-value = 0.0001352
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.908753 -4.234104
sample estimates:
mean of x mean of y
  6.00000  12.57143
```

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$



sort based on T-statistic

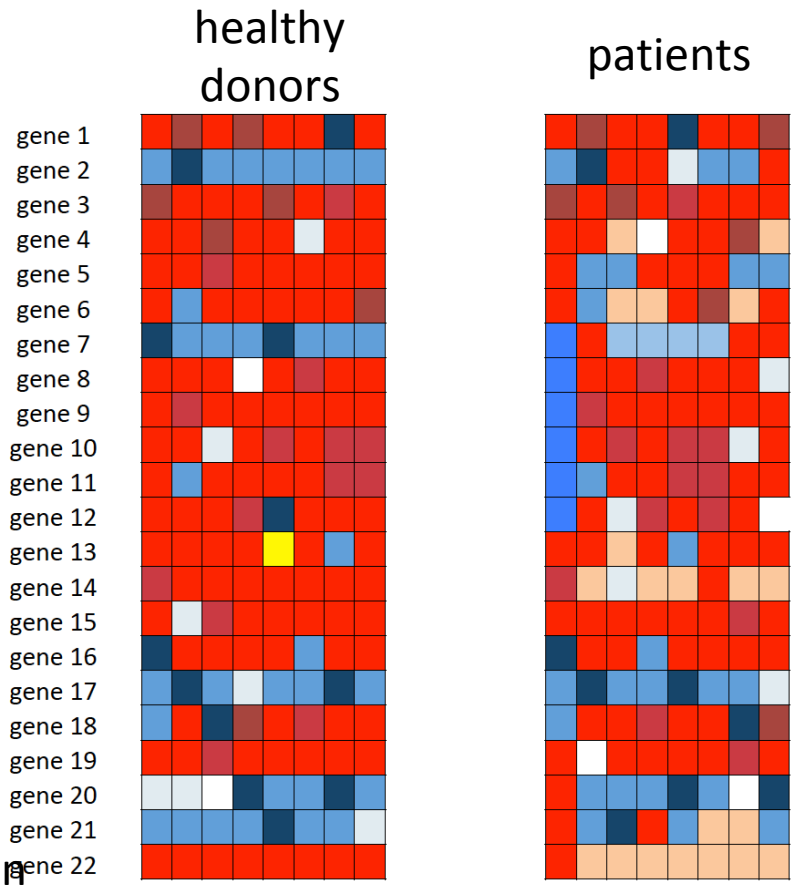
gene 13	-5
gene 17	-1
gene 20	-1
gene 1	0
gene 12	0
gene 15	0
gene 18	0
gene 19	0
gene 22	0
gene 3	0
gene 5	0
gene 8	0
gene 9	0
gene 10	0.4
gene 11	0.4
gene 16	0.4
gene 6	0.4
gene 21	0.6
gene 2	1
gene 7	1
gene 14	5
gene 4	5

Differential gene expression analysis using R

- Bioconductor

https://bioconductor.org/

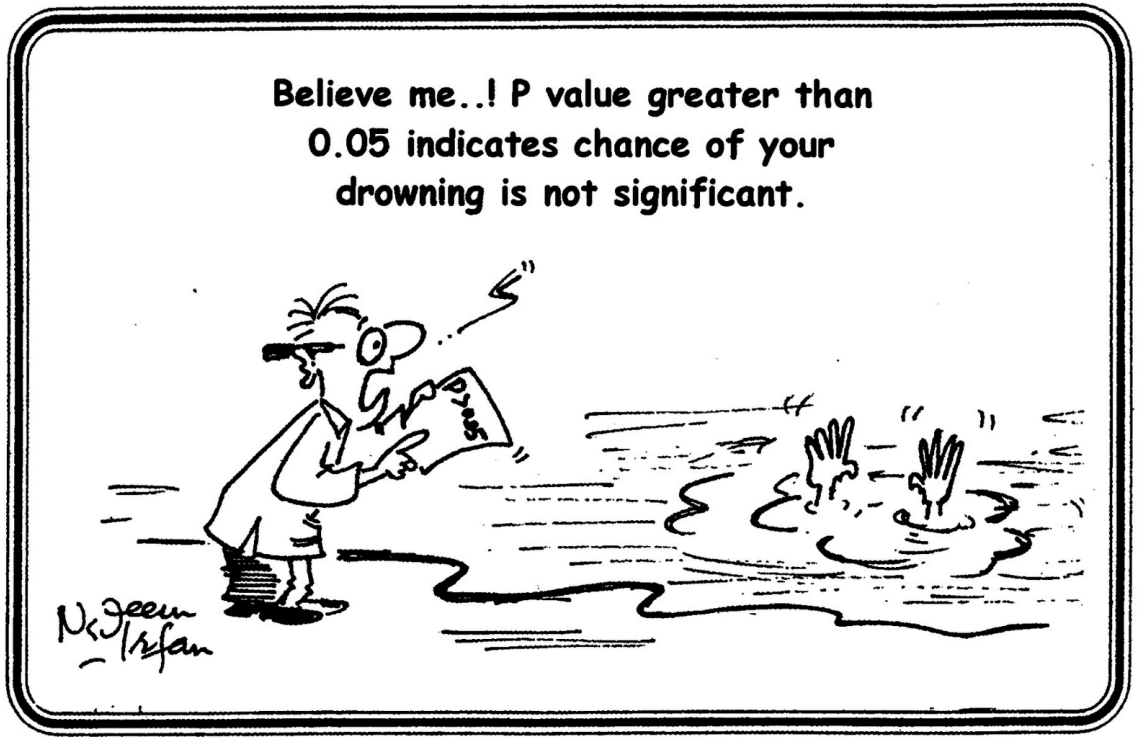
- Several packages :
  - limma: t-test
  - DESeq2: Wald test
  - edgeR: exact test



All allow for simultaneous statistical tests
for every gene, together with p-value calculation

What does p < 0.05 mean?

- It implies that it is acceptable to have a 5% probability to incorrectly reject the null hypothesis while it is correct.
- It means that if we repeat an experiment 20 times, we would reject the null hypothesis once because of random error.



Nadeem Irfan Bukhari
https://www.statsmadeeasy.net/2007/07/drowning-in-statistics/

P-value adjustment: what is it?

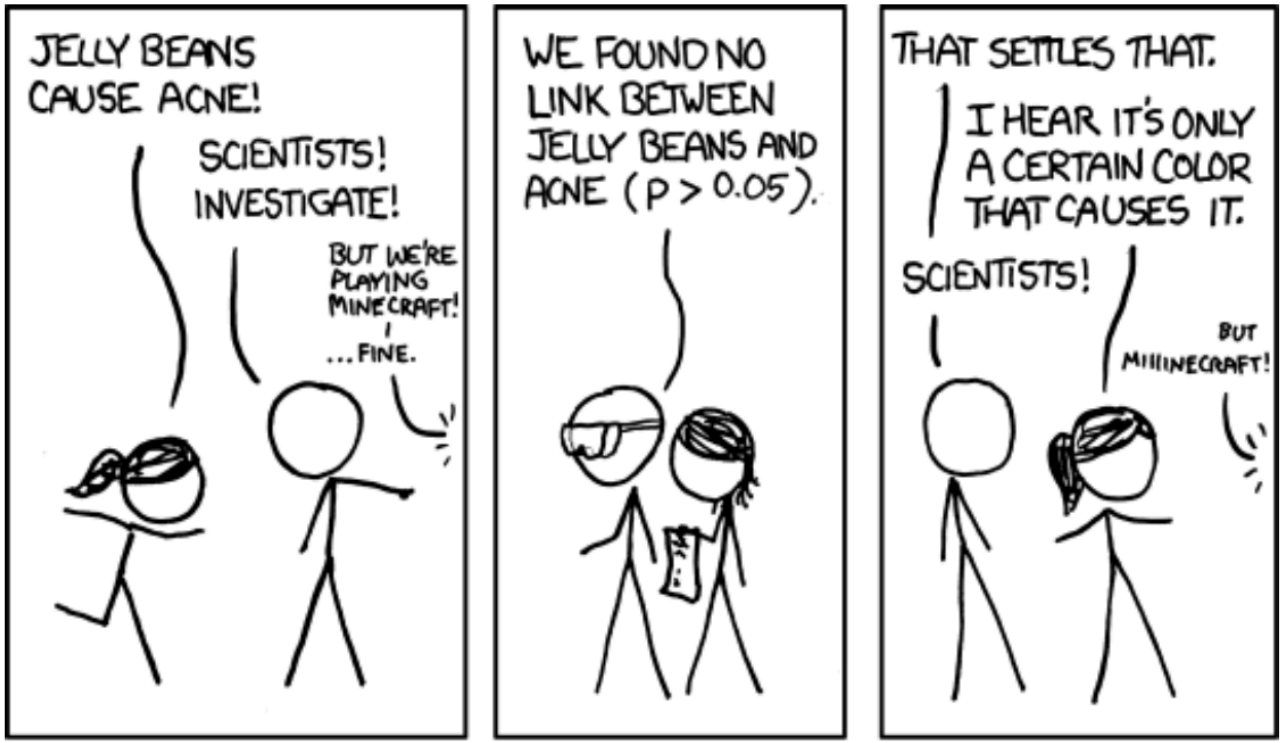
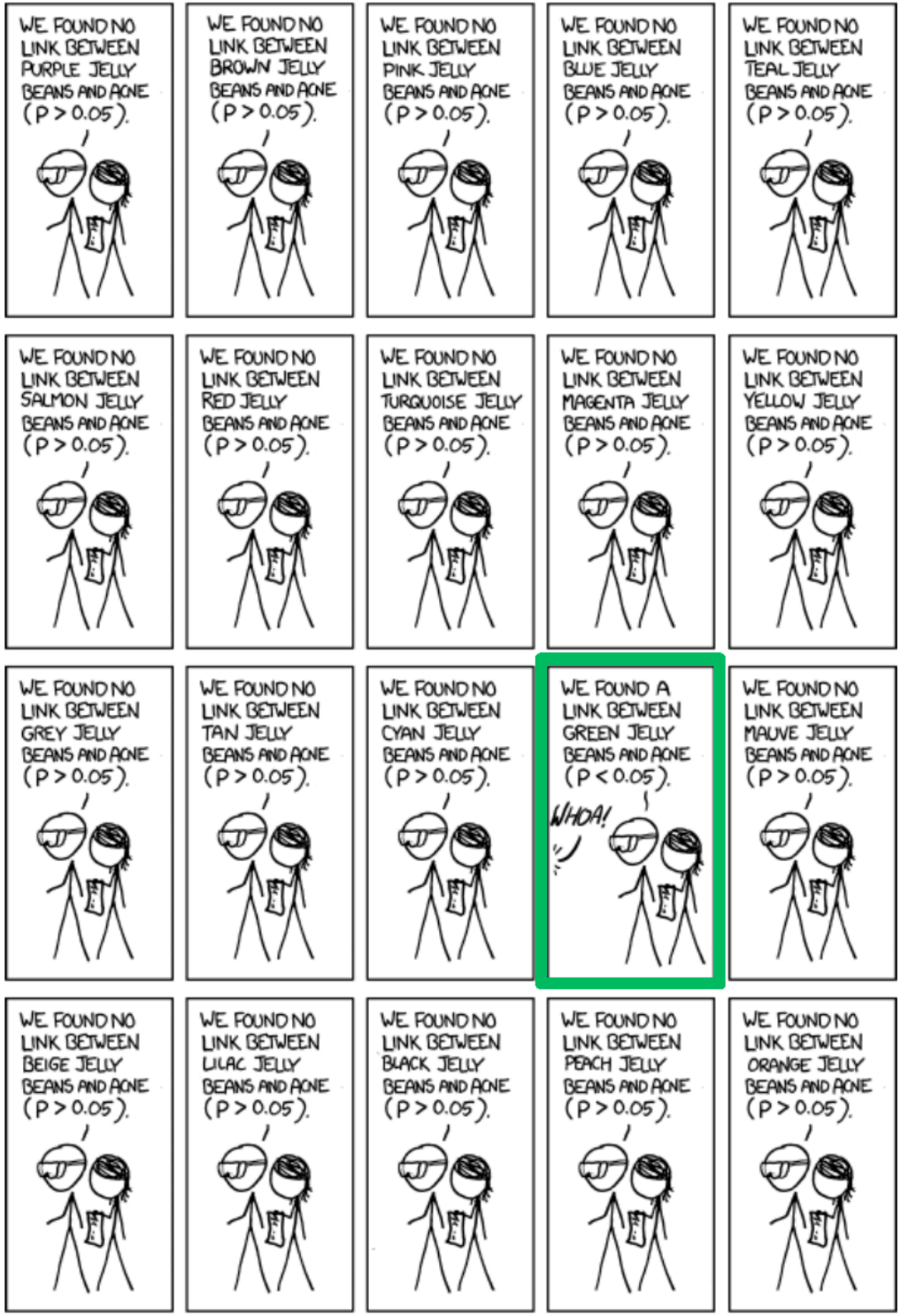
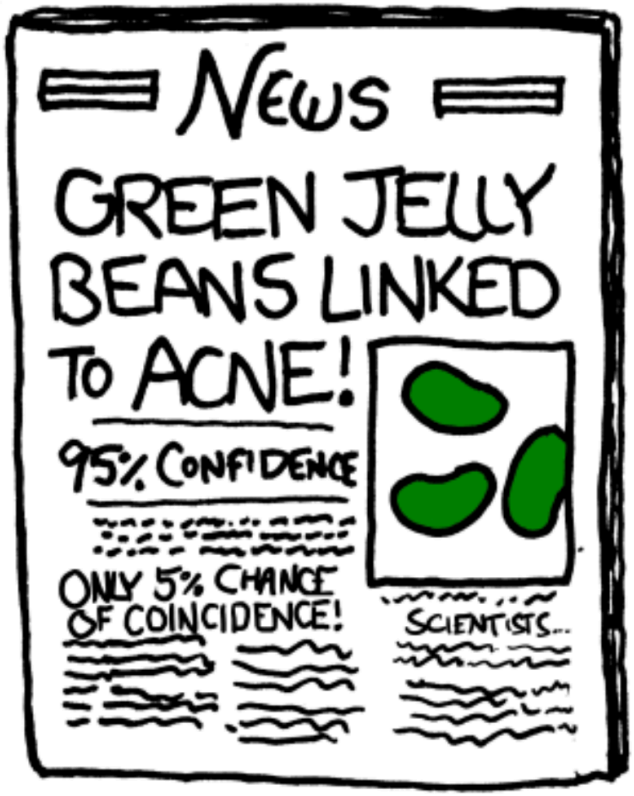


Photo by Patrick Fore on Unsplash

Cartoon: https://xkcd.com/882/
Paper on p-value adjustment: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6099145/


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).


News
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

Methods of p-value adjustment

- Bonferroni: the alpha level is divided by the total number of tests
- if we run k=20 tests:

0.05/k = 0.05/20=0.0025

Good for small number of tests but too conservative for thousands of genes

- Benjamini-Hochberg procedure (BH, decreases the FDR)
- Rank the p-values from smallest to largest, adjust less and less as the p-values get larger:

$p\text{-value}_1 * (n/1)$

$p\text{-value}_2 * (n/2)$

…

$p\text{-value}_k * (n/k) = p\text{-value}_k * 1$

n= total number of p-values (genes)

k= rank number of each p-value

https://www.youtube.com/watch?v=rZKa4tW2NKs