

Analysis of RNA-seq data

Day 3 recap



UNIVERSITY OF
CAMBRIDGE

Bioinformatics Training Facility

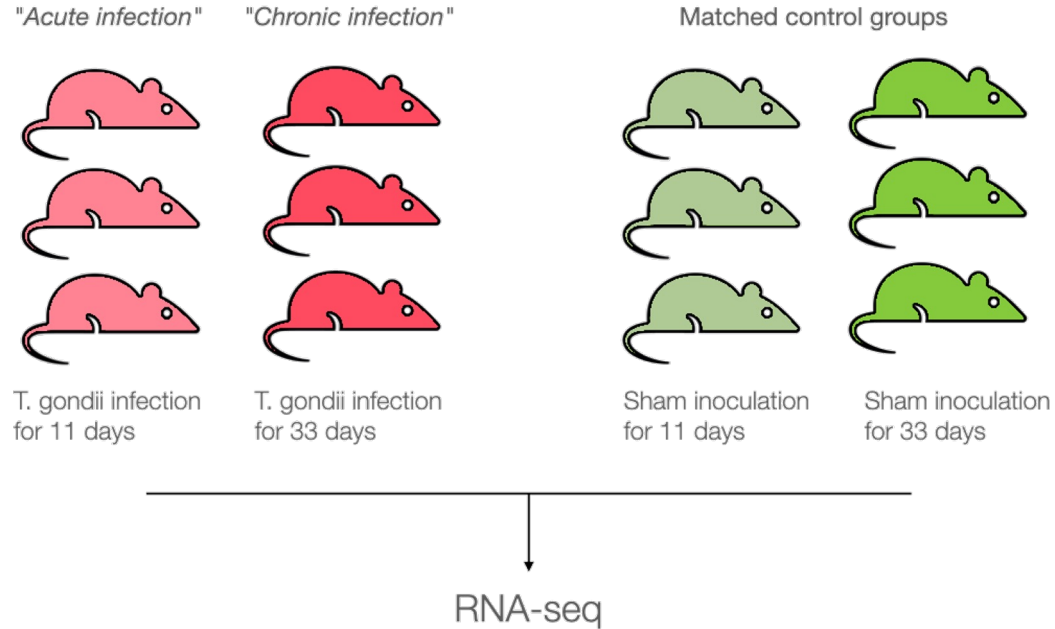


CANCER
RESEARCH
UK

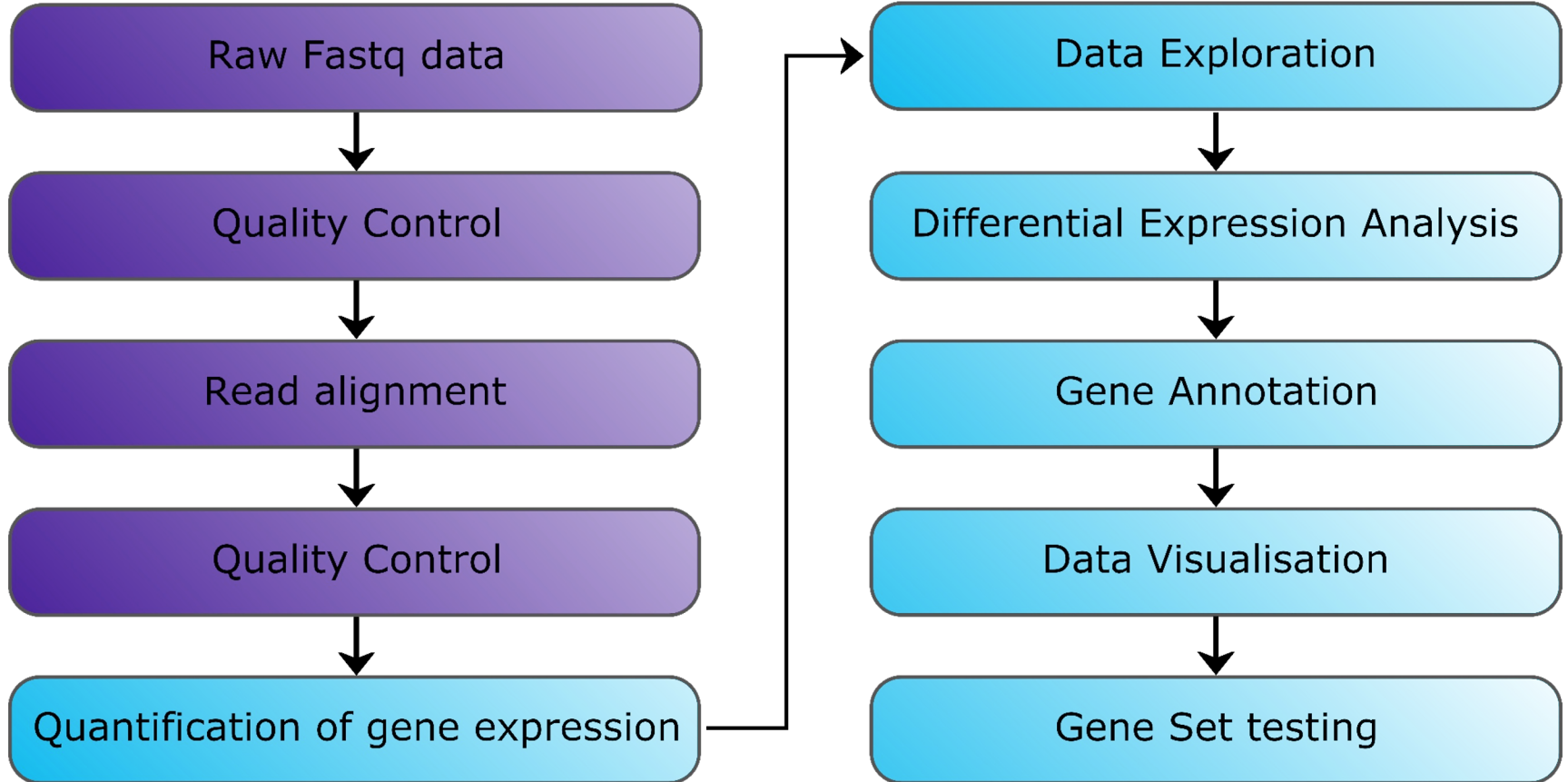
CAMBRIDGE
INSTITUTE

Transcriptomic Profiling of Mouse Brain During Acute and Chronic Infections by *Toxoplasma gondii* Oocysts

Rui-Si Hu^{1,2}, Jun-Jun He^{1*}, Hany M. Elsheikha³, Yang Zou¹, Muhammad Ehsan¹, Qiao-Ni Ma¹, Xing-Quan Zhu^{1,4} and Wei Cong^{5*}



Bioinformatics Analysis Workflow

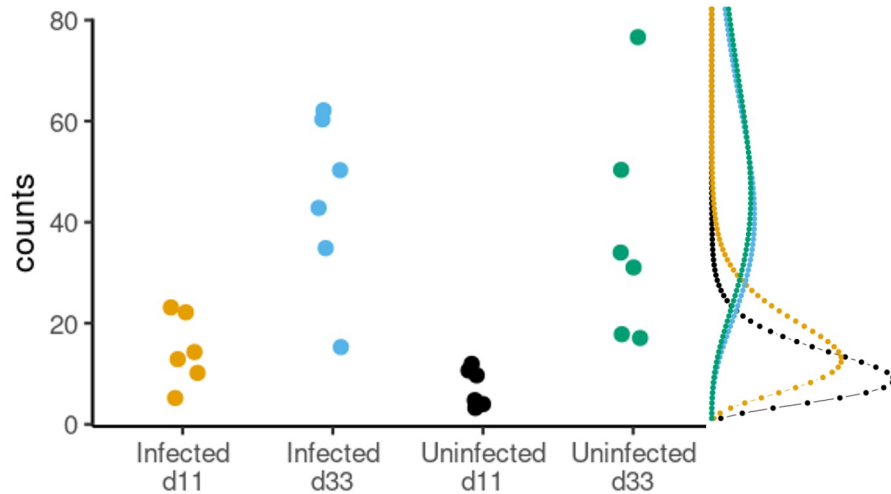


Linear Modelling

$$\text{Counts} \sim NB(\mu, \phi)$$

$$\mu = sq$$

$$\log_2(q) = \beta_0 + \beta_1 \times \text{Status} + \beta_2 \times \text{TimePoint}$$



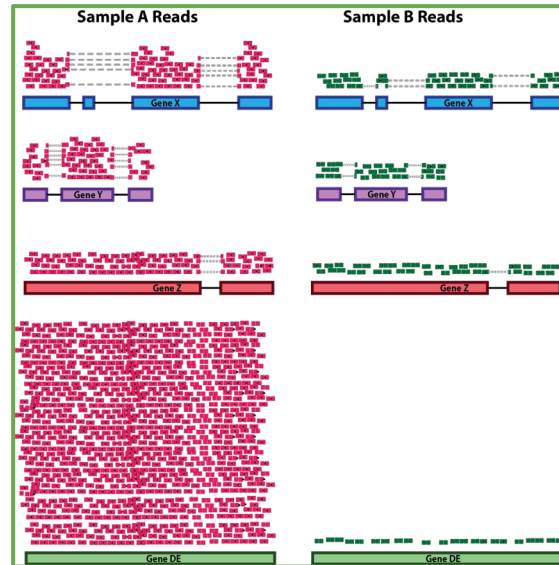
Linear Modelling

$$\text{Counts} \sim NB(\mu, \phi)$$

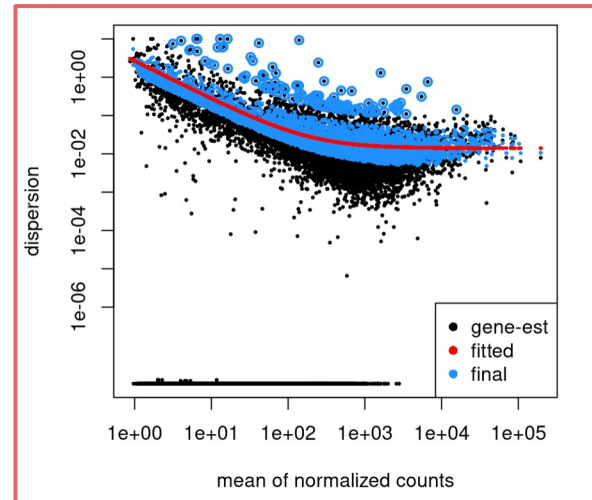
$$\mu = sq$$

$$\log_2(q) = \beta_0 + \beta_1 \times \text{Status} + \beta_2 \times \text{TimePoint}$$

Sample Scaling Factors



Gene Dispersion



DESeq2 | Models and Hypothesis Testing | One-Factor Model

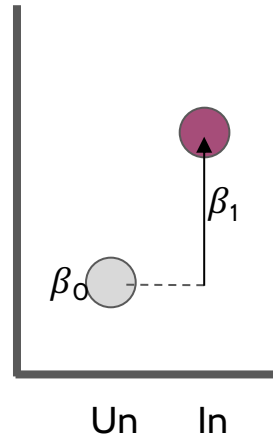
$$\log_2(q) = \beta_0 + \beta_1 \times Status$$

Formula syntax:

`~ Status`

Null hypothesis:

Infected vs Uninfected
 $\beta_1 = 0$



DESeq2 - Defining Models and Hypothesis Testing

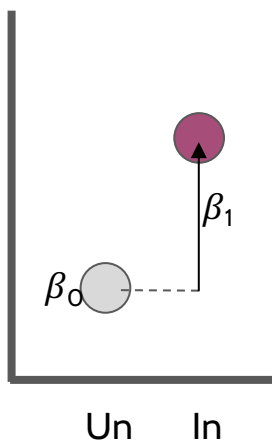
$$\log_2(q) = \beta_0 + \beta_1 \times Status$$

Formula syntax:

`~ Status`

Null hypothesis:

Infected vs Uninfected
 $\beta_1 = 0$



```
30
31 # Specify the model
32 simple.model <- as.formula(~ Status)
33
34 # create the DESeqDataSet object
35 ddsObj <- DESeqDataSetFromTximport(txi = txi,
36                                   colData = sampleinfo,
37                                   design = simple.model)
38
39 # Run the analysis
40 ddsObj <- DESeq(ddsObj)
41
42 # Extract results
43 results(ddsObj, alpha=0.05)
44
```

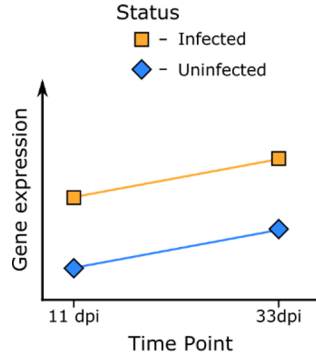
log2 fold change (MLE): Status Infected vs Uninfected

Wald test p-value: Status Infected vs Uninfected

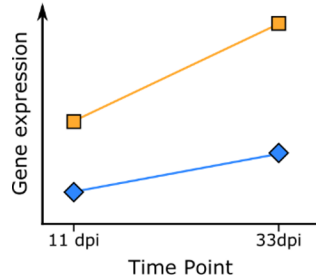
DataFrame with 10 rows and 6 columns

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
	<numeric>	<numeric>	<numeric>	<numeric>	<numeric>	<numeric>
ENSMUSG00000000001	1102.56094	-0.00802952	0.1028767	-0.078050	9.37788e-01	9.75584e-01
ENSMUSG00000000028	58.60055	0.30498077	0.2543119	1.199239	2.30435e-01	4.80598e-01
ENSMUSG00000000037	49.23586	-0.05272685	0.4168621	-0.126485	8.99348e-01	9.61314e-01
ENSMUSG00000000049	7.98789	0.38165132	0.6448695	0.591827	5.53966e-01	7.72123e-01
ENSMUSG00000000056	1981.00402	-0.16921845	0.1285416	-1.316449	1.88024e-01	4.26492e-01
ENSMUSG00000000058	606.92752	-0.17499624	0.1060284	-1.650466	9.88476e-02	2.86870e-01
ENSMUSG00000000078	1297.09024	0.50693550	0.0951759	5.326302	1.00232e-07	2.48994e-06
ENSMUSG00000000085	1063.43041	0.17636885	0.1346400	1.309929	1.90220e-01	4.29926e-01
ENSMUSG00000000088	2285.63159	-0.28670431	0.0878942	-3.261926	1.10658e-03	1.08587e-02
ENSMUSG00000000093	92.71580	0.07122976	0.2136598	0.333379	7.38848e-01	8.81258e-01

Two Factor Models



Additive model: gene expression changes with status and time point, but the difference in gene expression between infected and uninfected is the same irregardless of the time point



Interaction model: gene expression changes with status and time point, but the difference in gene expression between infected and uninfected changes depending on the time point

DESeq2 | Models and Hypothesis Testing | Two-Factor Additive Model

$$\log_2(q) = \beta_0 + \beta_1 \times Status + \beta_2 \times TimePoint$$

Formula syntax:

`~ Status + TimePoint`

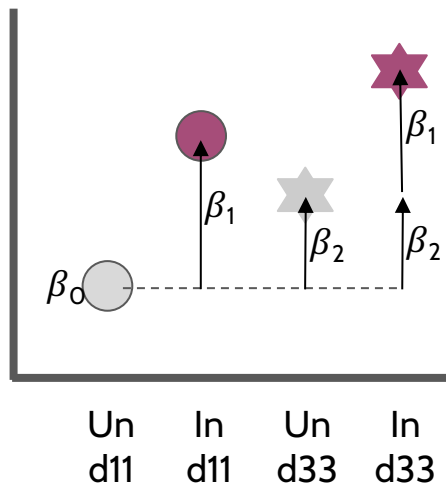
Null hypothesis:

Infected vs Uninfected

$$\beta_1 = 0$$

d33 vs d11

$$\beta_2 = 0$$



DESeq coefficient names:

$\beta_0 \rightarrow$ Intercept

$\beta_1 \rightarrow$

Status_Infected_vs_Uninfected

$\beta_2 \rightarrow$ TimePoint_d33_vs_d11

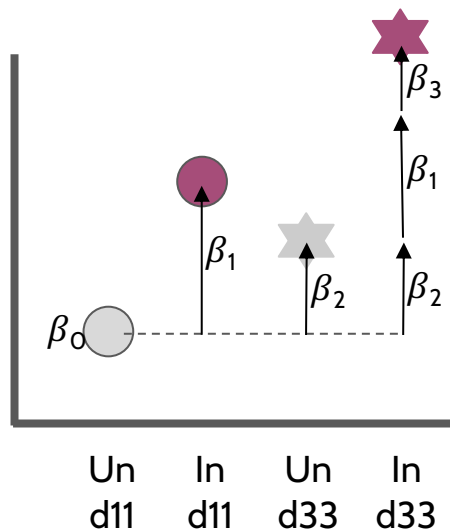
DESeq2 | Models and Hypothesis Testing | **Two-Factor Interaction Model**

$$\log_2(q) = \beta_0 + \beta_1 * Status + \beta_2 * TimePoint + \beta_3 * Status : TimePoint$$

Formula syntax:

`~ Status + TimePoint + Status:TimePoint`

`~ Status * TimePoint`



Null hypothesis:

Infected vs Uninfected (d11)

$$\beta_1 = 0$$

Infected vs Uninfected (d33)

$$\beta_1 + \beta_3 = 0$$

d33 vs d11 (Uninfected)

$$\beta_2 = 0$$

d33 vs d11 (Infected)

$$\beta_2 + \beta_3 = 0$$

Interaction (“Difference of differences”)

$$\beta_3 = 0$$

Model Specification in *DESeq2* | Likelihood-ratio Test

The default test in *DESeq2* is the Wald test, testing for null hypothesis that LFC = 0

And alternative is the **Likelihood Ratio Test**

$$LR = -2\ln\left(\frac{L(m_1)}{L(m_2)}\right)$$

~ genotype + timepoint + genotype:timepoint **VS** ~ genotype + timepoint

Example:

```
design(dds) <- ~ genotype + timepoint + genotype:timepoint  
  
dds <- DESeq(dds, test = "LRT",  
             reduced = ~ genotype)
```