

Analysis of RNA-seq data

Day 3 recap



UNIVERSITY OF
CAMBRIDGE

Bioinformatics Training Facility

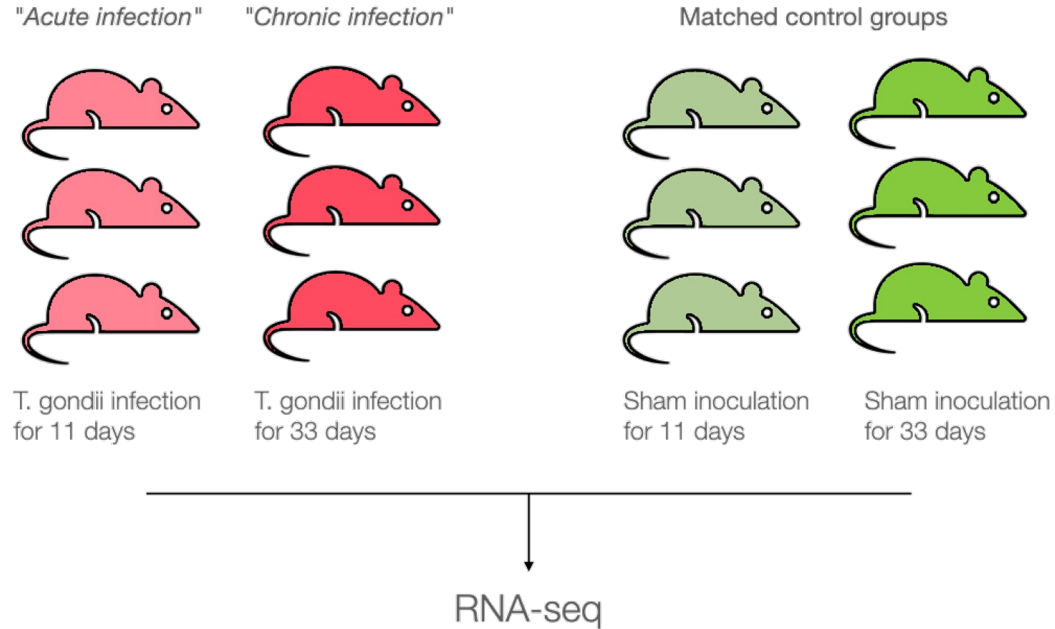


CANCER
RESEARCH
UK

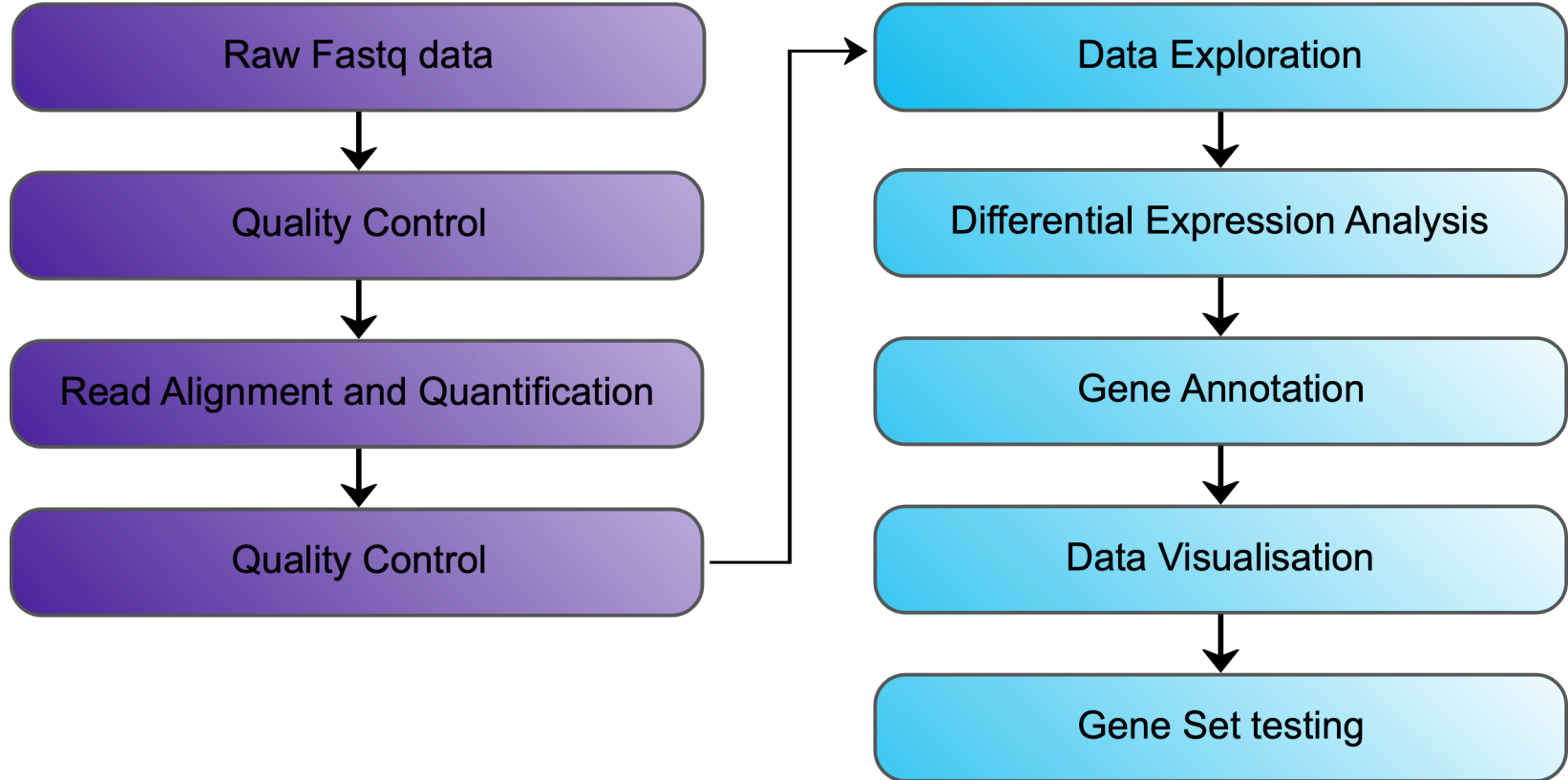
CAMBRIDGE
INSTITUTE

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



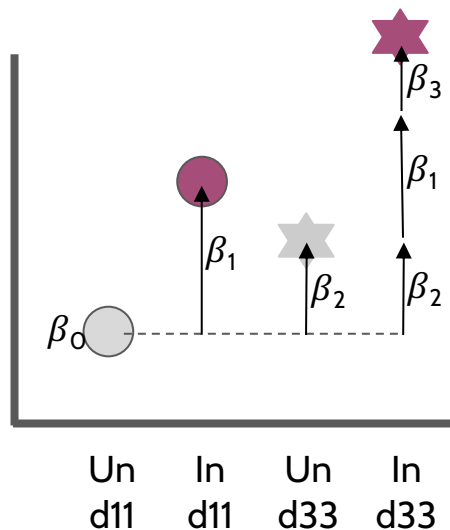
DESeq2 | Models and Hypothesis Testing | **Two-Factor Interaction Model**

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

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”)

$$\beta_3 = 0$$

DESeq2 | Models and Hypothesis Testing | Summary

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
	<numeric>	<numeric>	<numeric>	<numeric>	<numeric>	<numeric>
gene1	32.80405	0.359444	0.598072	0.601004	0.5478372	0.923764
gene2	4.01072	3.407763	1.649827	2.065527	0.0388732	0.641407
gene3	7.01837	0.743337	0.994100	0.747749	0.4546118	0.923764
gene4	1.51006	2.814822	2.464686	1.142061	0.2534287	0.923764
gene5	11.23166	0.480522	0.894709	0.537071	0.5912189	0.923764
...
gene96	16.21864	0.684962	0.809892	0.845745	0.3976952	0.923764
gene97	2.91349	1.784327	1.790046	0.996805	0.3188590	0.923764
gene98	13.29915	-0.634070	0.768728	-0.824830	0.4094680	0.923764
gene99	82.45653	-0.963147	0.505109	-1.906810	0.0565452	0.799710
gene100	6.25763	1.673078	1.252839	1.335429	0.1817359	0.923764

- Consider Experimental Design – Knowledge of the biology
- QC your analysis at every stage
- Select the simplest appropriate model for your data which best describes the majority of genes
- Know when to visit a bioinformatician or statistician.