



# Placental Bioinformatics Course

## Practical

Dr Russell S. Hamilton

Email: rsh46@cam.ac.uk  
Twitter: @drrshamilton

Dr Xiaohui Zhao

Email: xz289@cam.ac.uk

Dr Malwina Prater

Email: mn367@cam.ac.uk

Course Materials:

<https://github.com/CTR-BFX/2019-PlacentalBiologyCourse>

License:

Attribution-Non Commercial-Share Alike CC BY-NC-SA ( <https://creativecommons.org/licenses/by-nc-sa/> )

Attribution: You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NonCommercial: You may not use the material for commercial purposes.

ShareAlike: If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.



Version 0.1: 20190619

# Data Introduction

## RNA-Seq Differential Gene expression between the Placenta and Yolk Sac Mouse

- Data availability**

mRNA-Seq datasets have been deposited in Gene Expression Omnibus (GEO) under accession numbers GSE66137 and GSE66138. NanoString datasets have been deposited in GEO under accession number GSE79767.

### Bioinformatics Top Tip:

Download fastq files directly from  
<http://www.ebi.ac.uk/ena>



### Warning:

This is a demo with a reduced data set and parameters,  
so take any genes identified with caution

### RESEARCH ARTICLE

Deficiency of the placenta- and yolk sac-specific tristetraprolin family member ZFP36L3 identifies likely mRNA targets and an unexpected link to placental iron metabolism

Deborah J. Stumpo<sup>1</sup>, Carol S. Trempus<sup>2</sup>, Charles J. Tucker<sup>3</sup>, Weichun Huang<sup>4</sup>, Leping Li<sup>4</sup>, Kimberly Kluckman<sup>5</sup>, Donna M. Bortner<sup>3</sup> and Perry J. Blackshear<sup>1,6,\*</sup>

### ABSTRACT

The ZFP36L3 protein is a rodent-specific, placenta- and yolk sac-specific member of the tristetraprolin (TTP) family of CCHC tandem zinc finger proteins. These proteins bind to AU-rich elements in target mRNAs, and promote their deadenylation and decay. We addressed the hypotheses that the absence of ZFP36L3 would result in the accumulation of target transcripts in placenta and/or yolk sac, and that some of these would be important for female reproductive physiology and overall fecundity. Mice deficient in ZFP36L3 exhibited decreased neonatal survival rates, but no apparent morphological changes in the placenta or surviving offspring. We found *Zfp36l3* to be paternally imprinted, with profound parent-of-origin effects on gene expression. The protein was highly expressed in the syncytiotrophoblast cells of the labyrinth layer of the placenta, and the epithelial cells of the yolk sac. RNA-Seq of placental mRNA from *Zfp36l3* knockout (KO) mice revealed many significantly upregulated transcripts, whereas there were few changes in KO yolk sacs. Many of the upregulated placental transcripts exhibited decreased decay rates in differentiated trophoblast stem cells derived from KO blastocysts. Several dozen transcripts were deemed high probability targets of ZFP36L3; these include proteins known to be involved in trophoblast and placenta physiology. Type 1 transferrin receptor mRNA was unexpectedly decreased in KO placentas, despite an increase in its stability in KO stem cells. This receptor is crucial for placental iron uptake, and its decrease was accompanied by decreased iron stores in the KO fetus, suggesting that this intrauterine deficiency might have deleterious consequences in later life.

**KEY WORDS:** Deadenylation, mRNA binding proteins, mRNA decay, Placenta, Zinc finger proteins

### INTRODUCTION

The tristetraprolin (TTP) family of CCHC tandem zinc finger proteins consists of mRNA-binding proteins that are thought to regulate gene expression by binding to and promoting the decay of mRNAs containing specific types of AU-rich element-binding sites (Blackshear and Perera, 2014; Brooks and Blackshear, 2013).

<sup>1</sup>Laboratory of Signal Transduction, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA. <sup>2</sup>Laboratory of Clinical Research, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA. <sup>3</sup>Confocal Microscopy Core, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA. <sup>4</sup>Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA. <sup>5</sup>Translational Research Core, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA. <sup>6</sup>Department of Medicine and Biochemistry, Duke University Medical Center, Durham, NC 27710, USA.

\*Author for correspondence (blacks009@niehs.nih.gov)

Received 1 September 2015; Accepted 23 February 2016

These sites are generally located in the 3'-untranslated regions (3'UTRs) of mRNAs and have the optimal sequence of UUAUUUAU (Lai et al., 2005); physiologically relevant confirmed target transcripts often contain multiple binding sites in close proximity to each other (Carballo et al., 1998, 2000). The mechanism of the induced mRNA instability is not entirely understood, but appears to involve initial stimulation of deadenylation, or removal of the poly(A) tail, thought to be the rate-limiting step in mRNA decay in all eukaryotes (Carballo et al., 2000; Fabian et al., 2013; Lai et al., 2003).

In most mammals, the TTP protein family consists of three members, which behave similarly in biochemical assays of mRNA binding and destabilization (Frederick et al., 2008; Lai et al., 2003). However, knockout (KO) mice for these three family members exhibit dramatic physiological specificity. For example, TTP KO mice (*Zfp36*<sup>-/-</sup>) develop a systemic inflammatory syndrome that is largely due to chronic elevations in tumor necrosis factor alpha (TNF $\alpha$ ); the *Tyf* mRNA was found to be a direct target of TTP binding and induced mRNA destabilization (Carballo and Blackshear, 2001; Carballo et al., 1998; Taylor et al., 1996). KO of a second family member, *Zfp36l1*, results in embryonic lethality, apparently due to failure of chorioallantoic fusion, an essential step in the development of the umbilical circulation (Stumpo et al., 2004). The third family member, *Zfp36l2*, is crucial for definitive hematopoiesis (Stumpo et al., 2009).

A fourth mammalian gene in this family, *Zfp36l3*, is an X chromosome gene expressed only in the yolk sac and placenta of certain rodents, including mice and rats (Blackshear et al., 2005; Frederick et al., 2008; Gingerich et al., 2016). The ZFP36L3 protein differs from its other family members in several respects; for example, in contrast to the other family members, which are nucleocytoplasmic shuttling proteins (Phillips et al., 2002), the mouse and rat ZFP36L3 proteins contain long series of carboxyl-terminal repeats that serve to maintain the protein in the cytoplasm (Frederick et al., 2008). We describe here the development of a *Zfp36l3* KO mouse, and its phenotypic and biochemical characterization. Our findings highlight the importance of ZFP36L3 in mouse fertility, as well as its influence on post-transcriptional gene expression in placenta.

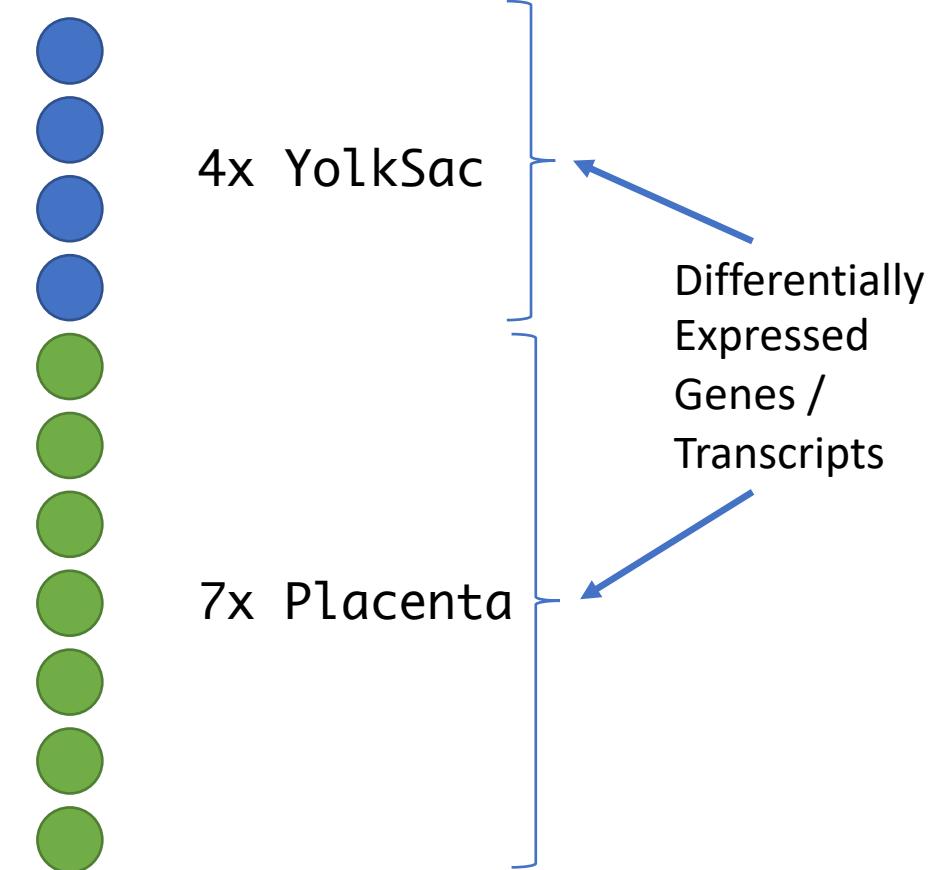
### RESULTS

#### Expression of ZFP36L3 during normal development

In placenta, *Zfp36l3* mRNA was readily detected by embryonic day (E) 9.5, increased to near maximum levels by E14.5, and then remained elevated until E18.5. In the yolk sac, *Zfp36l3* mRNA was essentially undetectable at E10.5, and then began accumulating, reaching a peak by E18.5 (Fig. 1A). Transcripts for the other family members were readily detectable and largely constant in both yolk

# Introduction

sample	condition
SRR1811706	WT Yolk Sac
SRR1811707	WT Yolk Sac
SRR1811708	WT Yolk Sac
SRR1811709	WT Yolk Sac
SRR1823638	WT Placenta
SRR1823639	WT Placenta
SRR1823640	WT Placenta
SRR1823641	WT Placenta
SRR1823642	WT Placenta
SRR1823643	WT Placenta
SRR1823644	WT Placenta





# Bioinformatics Pipeline

Sequencing Files (FASTQ)

Perform quality control (adapter contamination, base quality)

Align reads to the genome/transcriptome

Summarise QC and alignment metrics

**FastQC**

**trim\_galore**

**kallisto**

**MultiQC**

**Command Line**  
(R/R-Studio, system call)

**Break**

**Lecture 2**

Perform differential gene/transcript expression analysis

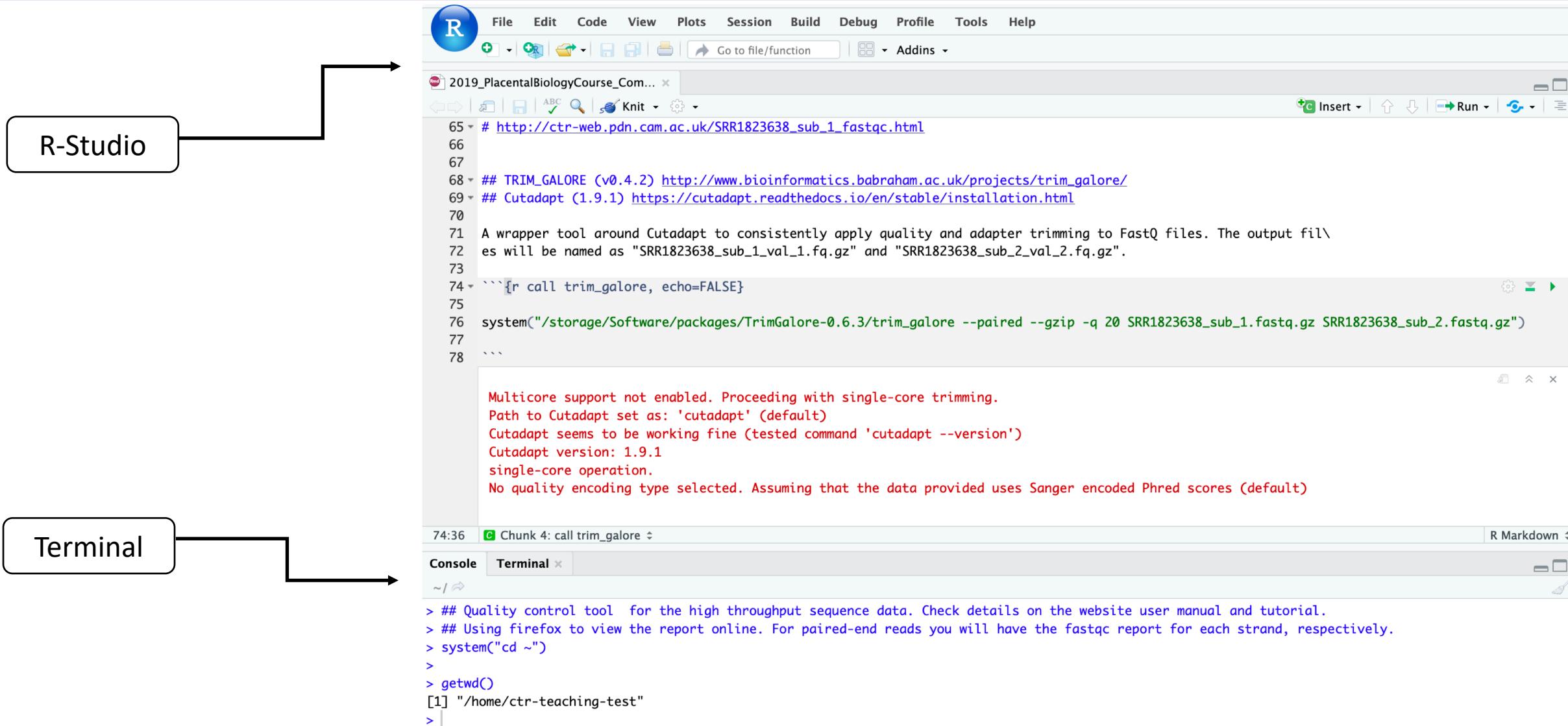
Perform Gene Ontology analysis

**DESeq2**

**clusterProfiler**

**R/R-Studio**

# Using the Bioinformatics Training Facility Computers



# Files Summary

## Course Materials

2019\_PlacentalBiologyCourse\_Lecture\_1.pptx  
2019\_PlacentalBiologyCourse\_Practical\_1.pptx  
**2019\_PlacentalBiologyCourse\_Commands\_Rversion.Rmd**  
2019\_PlacentalBiologyCourse\_Lecture\_2.pptx  
2019\_PlacentalBiologyCourse\_DESeq2.Rmd

## Publication

stumpo\_2016\_Development.pdf

## Sample Data

SRR1811706\_ES610\_WT\_Yolk\_Sac/  
SRR1811707\_ES611\_WT\_Yolk\_Sac/  
SRR1811708\_ES612\_WT\_Yolk\_Sac/  
SRR1811709\_ES613\_WT\_Yolk\_Sac/  
SRR1823638\_ES51\_WT\_Placenta/  
SRR1823639\_ES51\_WT\_Placenta/  
SRR1823640\_ES52\_WT\_Placenta/  
SRR1823641\_ES52\_WT\_Placenta/  
SRR1823642\_ES53\_WT\_Placenta/  
SRR1823643\_ES54\_WT\_Placenta/  
SRR1823644\_ES55\_WT\_Placenta/

## Reference Genome

ENST\_ENSG\_GeneName.GRCm38.kallisto.table  
Mus\_musculus.GRCm38.cdna.all.idx

## Practical Data Sets

**SRR1823638\_sub\_1.fastq.gz**  
**SRR1823638\_sub\_2.fastq.gz**

### SRR1823638 Sequencing Data

SRR1823638\_1.fastq.gz\_trimming\_report.txt  
SRR1823638\_1\_fastqc.html  
SRR1823638\_1\_fastqc.zip  
SRR1823638\_1\_val\_1.fq.gz\_kallisto\_output/  
SRR1823638\_1\_val\_1\_fastqc.html  
SRR1823638\_1\_val\_1\_fastqc.zip  
SRR1823638\_2.fastq.gz\_trimming\_report.txt  
SRR1823638\_2\_fastqc.html  
SRR1823638\_2\_fastqc.zip  
SRR1823638\_2\_val\_2\_fastqc.html  
SRR1823638\_2\_val\_2\_fastqc.zip

### Kallisto Output

- abundance.h5
- abundance.tsv
- run\_info.json

Subset of real data for this practical 1M reads only (~10%)



# FastQC

<b>FastQC</b>	A quality control tool for high throughput sequence data
<b>Version</b>	0.11.5
<b>Download</b>	<a href="http://www.bioinformatics.babraham.ac.uk/projects/fastqc/">http://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>

**Terminal:**

Read 1

Read 2

```
$ fastqc SRR1823638_sub_1.fastq.gz SRR1823638_sub_2.fastq.gz
```

**Output:** (Link: [http://ctr-web.pdn.cam.ac.uk/SRR1823638\\_sub\\_1\\_fastqc.html](http://ctr-web.pdn.cam.ac.uk/SRR1823638_sub_1_fastqc.html))

*HTML Reports*

SRR1823638\_sub\_1\_fastqc.html  
SRR1823638\_sub\_2\_fastqc.html

*Archive of data/images*

SRR1823638\_sub\_1\_fastqc.zip  
SRR1823638\_sub\_2\_fastqc.zip

*Bioinformatics Top Tip:* Simon Andrews' <https://sequencing.qcfail.com/>



# trim\_galore

<b>trim_galore</b>	A wrapper tool around Cutadapt to consistently apply quality and adapter trimming to FastQ files
<b>Version</b>	0.6.3
<b>Download</b>	<a href="http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/">http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/</a>

## Terminal:

```
$ trim_galore --paired --gzip -q 20 SRR1823638_sub_1.fastq.gz SRR1823638_sub_2.fastq.gz
```

Treat as paired-end      Quality score threshold (PHRED > 20), default

## **Output:**

## *Trimmed Fastq files*

SRR1823638\_sub\_1\_val\_1.fq.gz  
SRR1823638\_sub\_2\_val\_2.fq.gz

# kallisto

**kallisto** Program for quantifying abundances of transcripts from RNA-Seq data, without the need for alignment  
**Version** 0.44.0  
**Download** <https://pachterlab.github.io/kallisto>

## Terminal:

```
$ kallisto quant -b 25 -i Mus_musculus.GRCm38.cdna.all.idx  
      -o kallisto_output SRR1823638_sub_1_val_1.fq.gz SRR1823638_sub_2_val_2.fq.gz
```

Number of bootstraps      Indexed transcriptome  
Output directory      Trimmed Read 1      Trimmed Read 2

## Output:

*Note command must be all on one single line*

**kallisto\_output/**  
**abundance.h5**  
**abundance.tsv**  
**run\_info.json**

# RNA-Seq Mapping Metrics: Counts Vs FPKM Vs TPM

## Counts

The number of reads mapping to a transcript or gene

Longer transcripts will generally have more mapped reads

- None of these are for comparing across samples
- Sample normalisation required as performed by DESeq2
- Kallisto output with counts and TPM

## FPKM (Fragments Per Kilobase of transcript per Million mapped reads)

Normalises the counts for the length of the transcript

## TPM (Transcripts Per Million)

Measurement of the proportion of transcripts in your pool of RNA

# MultiQC

**MultiQC**

Aggregate results from bioinformatics analyses across many samples into a single report

**Version**

1.1dev

**Download**

<http://multiqc.info/>

**Terminal:**

```
$ multiqc -f -i "Placental Biology Course 2018"  
    --filename "PlacentalBiologyCourse.multiqc_report.html" .
```

Overwrite existing report      A title for your report

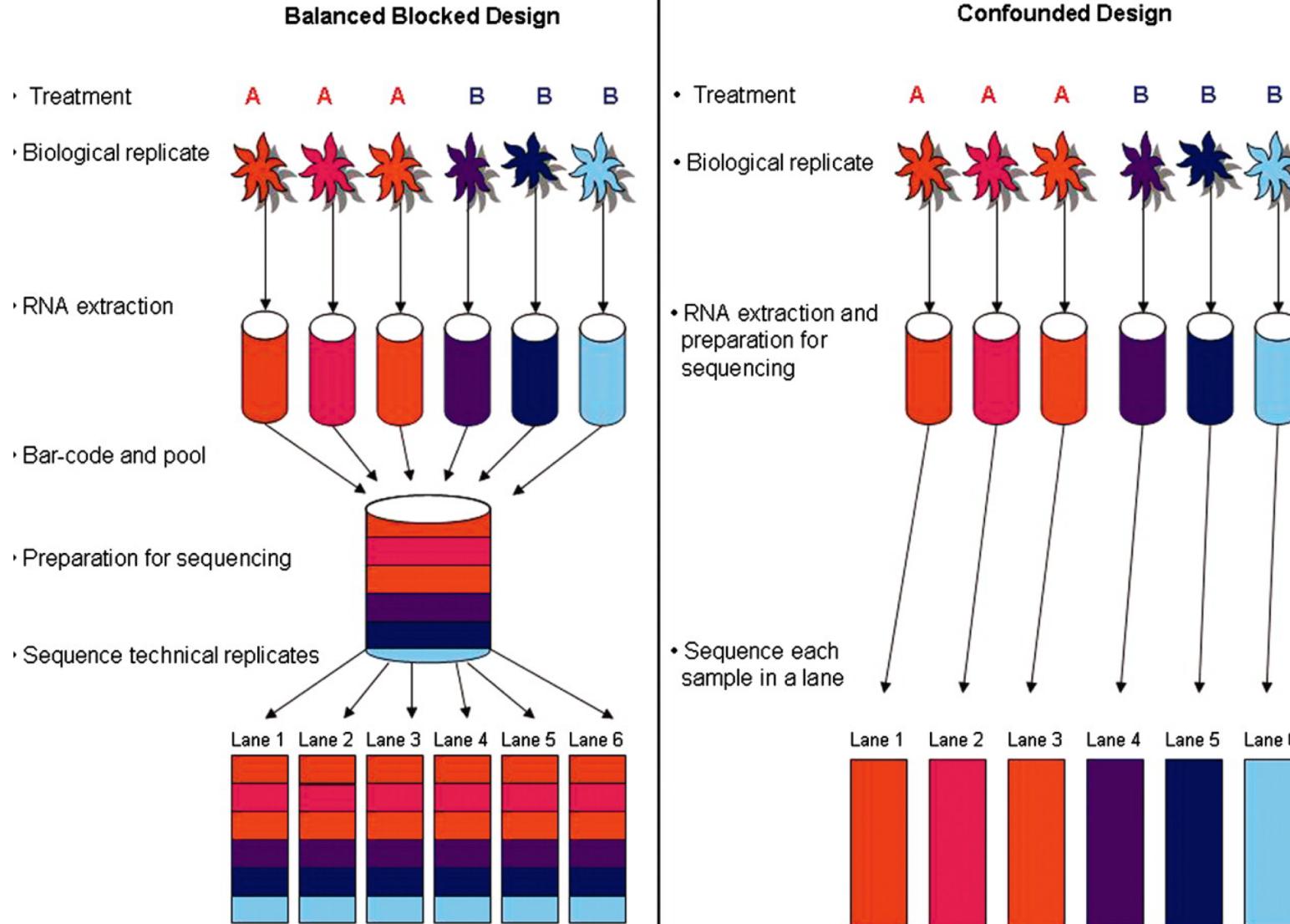
Output filename        
"." Is a special Linux symbol which means the current directory

**Output:** (link: <http://ctr-web.pdn.cam.ac.uk/multiqc.html>)

*HTML Report*

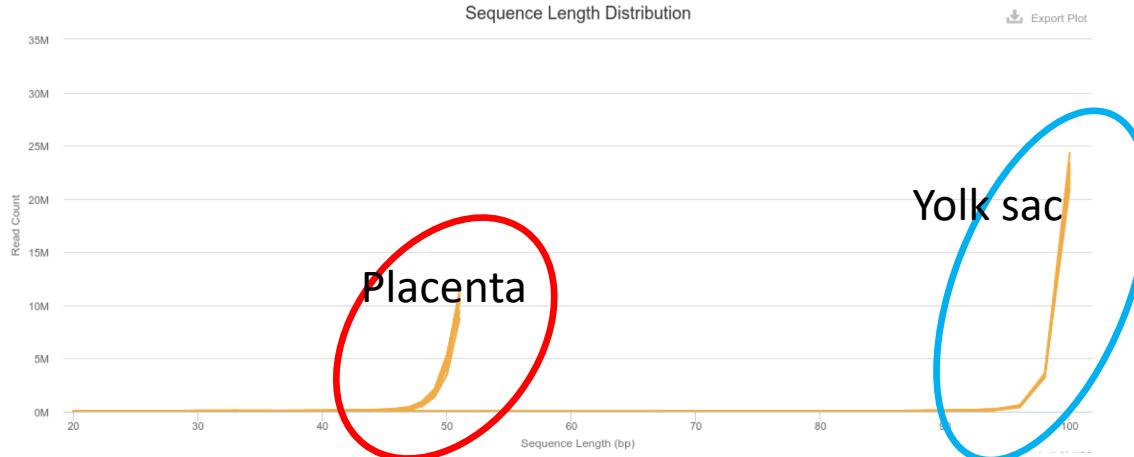
PlacentalBiologyCourse.multiqc\_report.html  
PlacentalBiologyCourse.multiqc\_report\_data

# Experimental Design



# QC Fastq Files----- Batch Effects

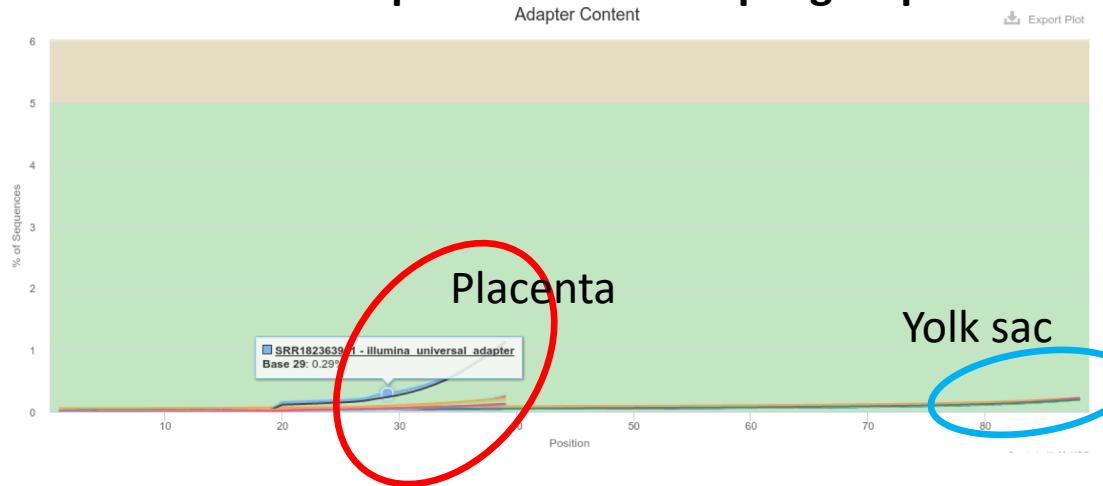
Sample groups have different read lengths



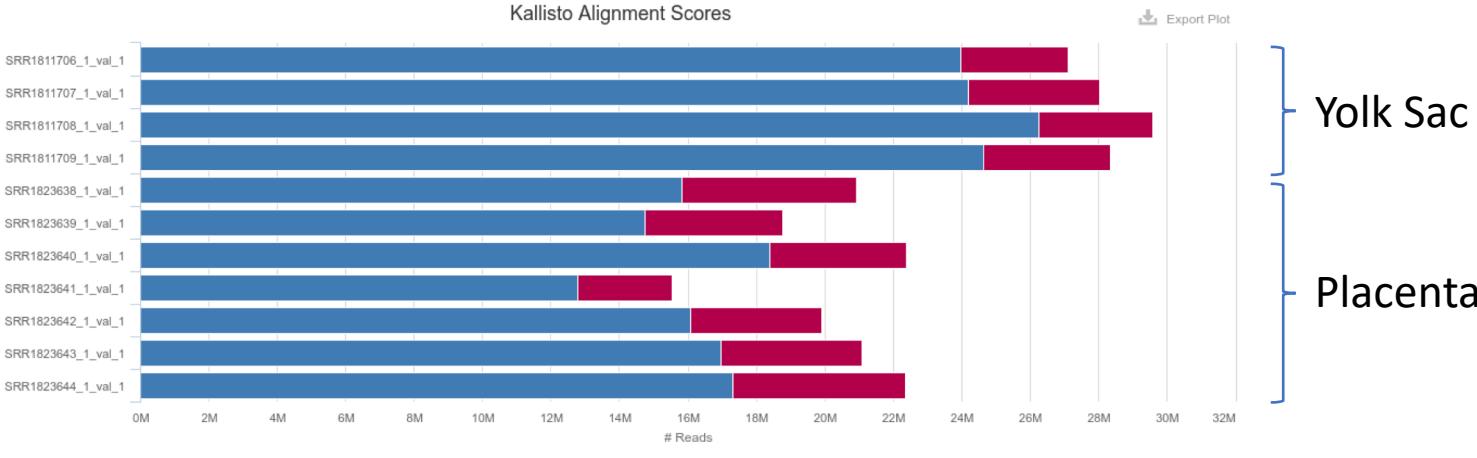
Some Placenta samples have low quality scores



There are adapters in both sample groups

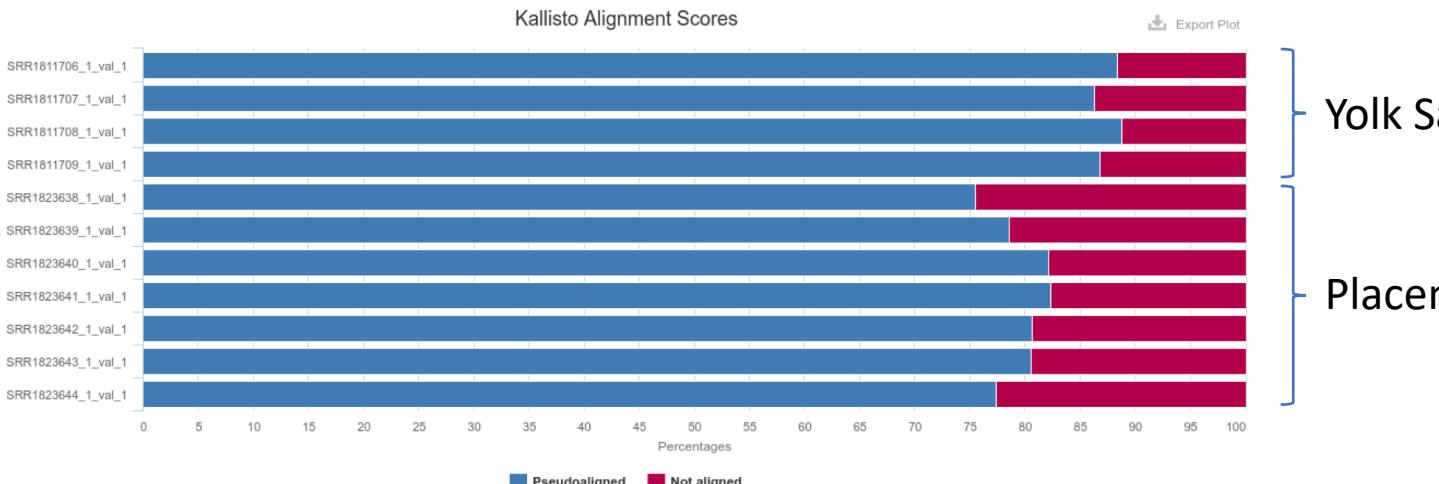


# QC Alignments



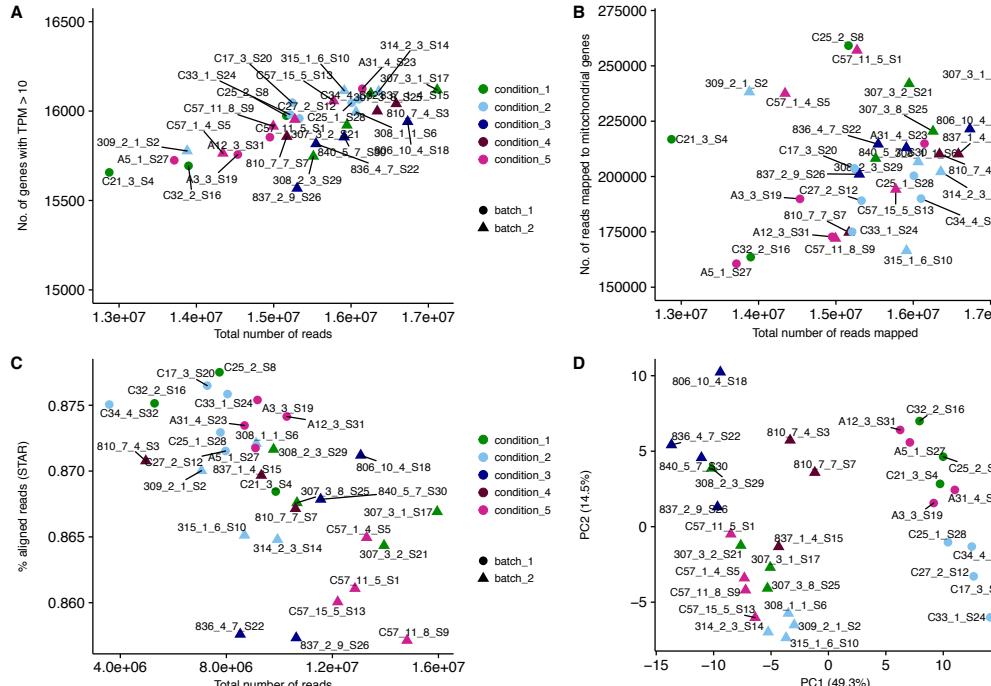
*Why do you never see 100% alignment?*

- Incomplete reference genomes / transcriptomes
- Repetitive reads hard to map uniquely



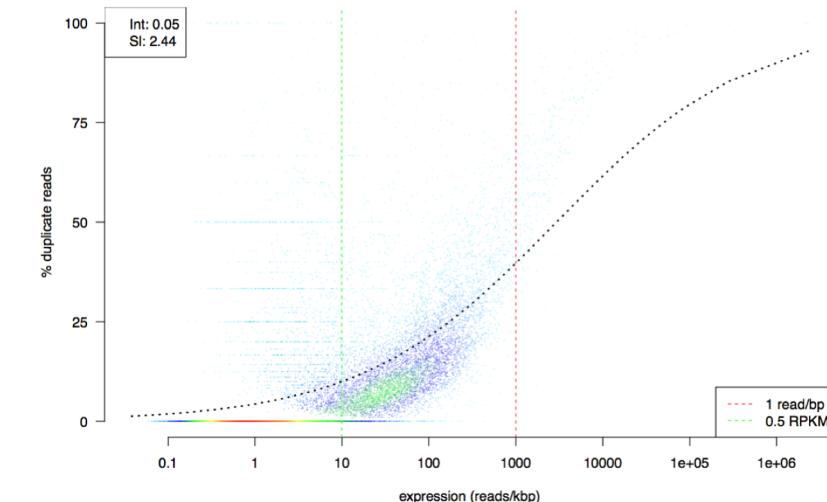
# Further QC

## rRNA and mt content



## dupRadar

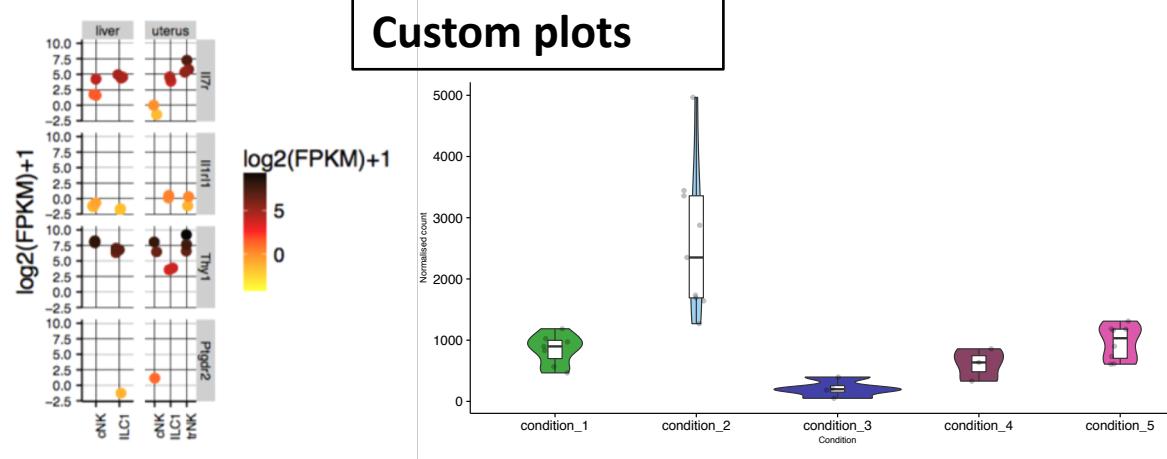
EXPERIMENT :: DupRadar scatter



## Batch effect correction

- e.g. Two sequencing experiments performed on different days by two different people
- Batch effect removal: Limma / Combat
- Use batch in comparison in DESeq2 design  
~ batch + condition

## Custom plots



# What is the next?

**Break**

**Lecture 2**

**Practical 2**