



Placental Bioinformatics Course

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

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.

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

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 0.3: 20180705



An introduction to RNA-Seq

Objective for the practical

To learn about how to process, analyse and interpret RNA-Seq data

Learning goals:

Experience using the Linux operating system (Ubuntu)

Learn how to run bioinformatics tools on the command line

Learn how to use R and R-studio

Visualise sequencing alignments in a genome browser

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. Stumpo¹, Carol S. Trempus², Charles J. Tucker², Weichuan Huang³, Leping Li⁴, Kimberly Kluckman⁵, Donna M. Bortner² and Perry J. Blackshear^{1,6,*}

ABSTRACT

The ZFP36L3 protein is a rodent-specific, placenta- and yolk sac-specific member of the tristetraprolin (TTP) family of CCCH tandem zinc-finger proteins. These proteins bind to AU-rich elements in target mRNAs, and promote their deadenylation and decay. We addressed the hypothesis 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 mastocytes. Several dozen transcripts were deemed high probability targets of ZFP36L3; these include proteins known to be involved in implantation and placenta physiology. Type 1 transferrin receptor mRNA was unexpectedly decreased in KO placenta, 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 CCCH 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 Preys, 2014; Brooks and Blackshear, 2013).

¹Laboratory of Signal Transduction, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA; ²Laboratory of Clinical Pharmacology, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA; ³Coastal Microscopy Core, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA; ⁴Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA; ⁵Theravance, Inc., Research Triangle Park, NC 27709, USA; ⁶Department of Molecular and Biochemical Medicine, Duke University Medical Center, Durham, NC 27710, USA

*Author for correspondence: (dev130389@nremsa.nih.gov)

Received 1 September 2015; Accepted 23 February 2016

1424

These sites are generally located in the 3'-untranslated regions (3'UTRs) of mRNAs and have the optimal sequence of UUAUAUUUU (Lai et al., 2009); physiologically relevant confirmed target transcripts often contain multiple binding sites in close proximity to each other (Carballo et al., 1998, 2008). 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., 2008; Fabian et al., 2013; Lai et al., 2009).

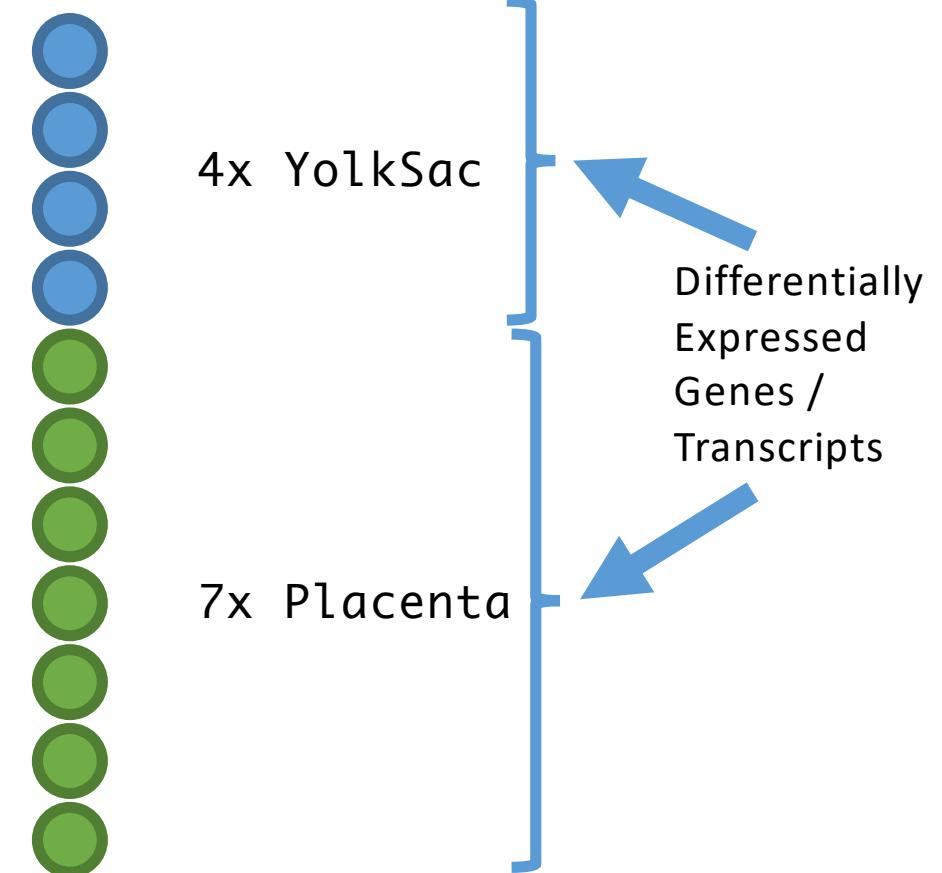
In most mammals, the TTP protein family consists of three members, which behave similarly in biochemical assays of mRNA binding and destabilization (Froderick et al., 2008; Lai et al., 2009). However, knockout (KO) mice for these three family members exhibit dramatic physiological specificity. For example, TTP KO mice (*Zfp36l3*^{-/-}) develop a systemic inflammatory syndrome that is largely due to chronic elevation in tumor necrosis factor alpha (TNFα); the *Puf* 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 choriocarcinoma fusion, an essential step in the development of the umbilical circulation (Stumpo et al., 2006). 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; Froderick et al., 2008; Gingrich et al., 2014). The ZFP36L3 protein differs from its other family members in several respects; for example, in contrast to the other family members, which are macrocytovironetting proteins (Phillips et al., 2002), the mouse and rat ZFP36L3 proteins contain long series of carboxy-terminal repeats that serve to maintain the protein in the cytoplasm (Froderick et al., 2008). We describe here the development of a *Zfp36l3* KO mouse, and its phenotype 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)

FastQC



Perform quality control (adapter contamination, base quality)

trim_galore

Align reads to the genome/transcriptome

kallisto

Command Line
&
FireFox

Summarise QC and alignment metrics

MultiQC

Coffee break

Perform differential gene/transcript expression analysis

sleuth/DESeq2

R/R-Studio

Visualise sequencing alignments in a genome browser

IGV

IGV

Look at differentially enriched genes

ensEMBL/DAVID

Firefox



Using the Bioinformatics Training Facility Computers

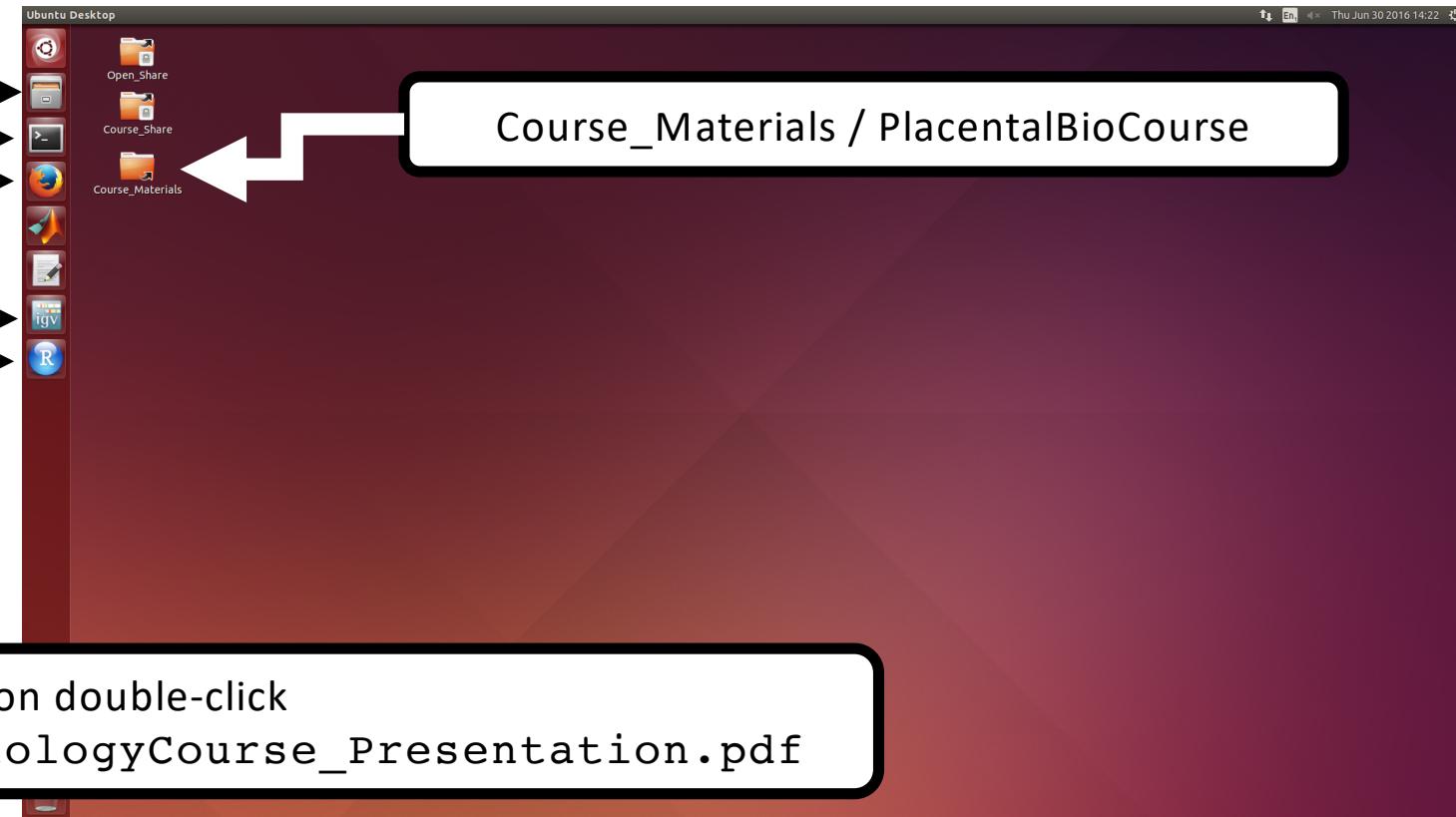
Finder / Windows Explorer

Terminal

Firefox

IGV

R-Studio



Linux :::: Ubuntu

Course Materials

2018_PlacentalBiologyCourse.pptx
2018_PlacentalBiologyCourse.pdf
2018_PlacentalBiologyCourse_Sleuth.Rmd
2018_PlacentalBiologyCourse_commands.txt

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/

Alignments (Premade)

IGV_Alignments/
ENST_ENSG_GeneName.GRCm38.kallisto.table
Mus_musculus.GRCm38.cdna.all.idx

Reference Genome

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

Placenta_SRR1823638_kallisto.bam
Placenta_SRR1823638_kallisto.bam.bai
YolkSac_SRR1811706_kallisto.bam
YolkSac_SRR1811706_kallisto.bam.bai

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



Using the Bioinformatics Training Facility Computers

Terminal

```
rhamilton — participant@winfields: ~/Course_Materials — ssh participant@131.111.146.51 -Y — 84x15
rsh46@login-sand3:~/scratch/CTR... participant@winfields: ~/Course_Materials...
participant@winfields:~$ cd Course_Materials
[participont@winfields:~/Course_Materials]
participant@winfields:~/Course_Materials$ cd PlacentalBiologyCourse
```

change directory Directory name

\$ cd Course_Materials

\$ cd PlacentalBioCourse

Bioinformatics Top Tip: More Linux Commands

| | |
|---------|--|
| ls | list files in directory |
| cd ~ | change back to home directory |
| tree | view files and directories in a hierarchical structure |
| history | view a list of the most recent commands used |

Caution!

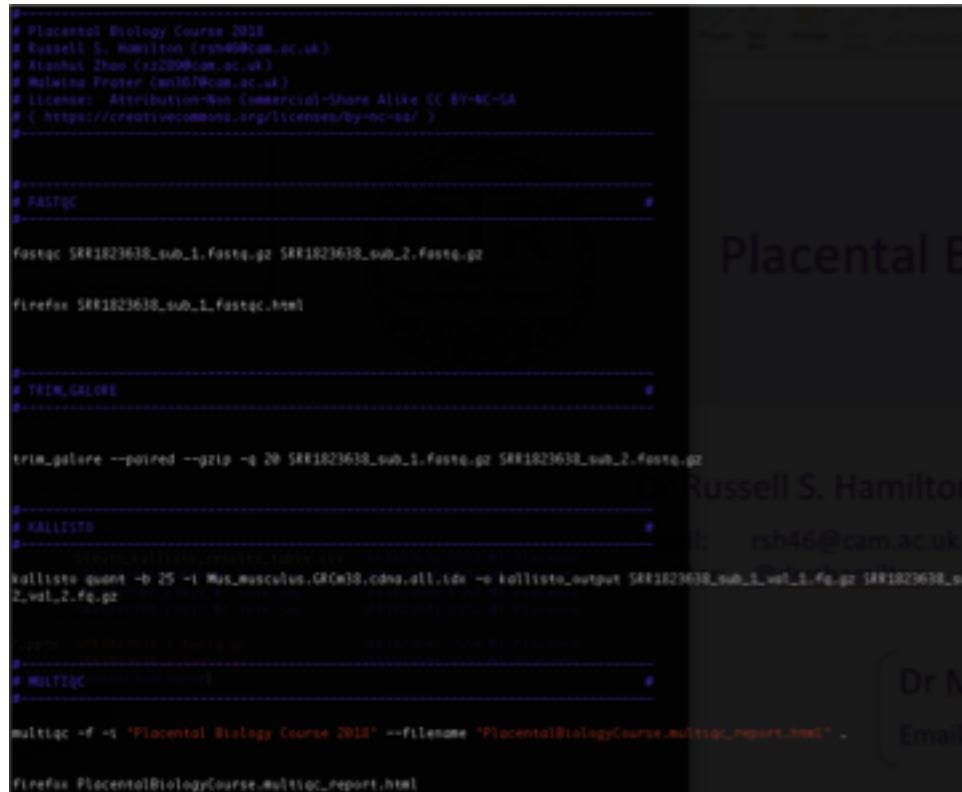
Commands are case sensitive

Take care to correctly specify spaces and flags (dashes)

Command Line

Copy & Paste from PDF or Powerpoint to the command line doesn't always work as expected

2018_PlacentalBiologyCourse_Commands.txt



```
# Placental Biology Course 2018
# Russell S. Hamilton (rsh46@cam.ac.uk)
# Xiaohui Zhao (xz289@cam.ac.uk)
# Malwina Prater (mn367@cam.ac.uk)
# License: Attribution-Non Commercial-Share Alike CC BY-NC-SA
# < https://creativecommons.org/licenses/by-nc-sa/ >
# 

# FASTQC
# 

fastqc SRR1823638_sub_1.fastq.gz SRR1823638_sub_2.fastq.gz

firefox SRR1823638_sub_1.fastqc.html

# 

# TRIM_GALORE
# 

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

# 

# KALLISTO
# 

kallisto quant -b 25 -l Mus_musculus.GRCm38.cdna.all.tcr -o kallisto_output SRR1823638_sub_1.vol_1.fq.gz SRR1823638_sub_2.vol_2.fq.gz

# 

# MULTICQ
# 

multiqc -f -s "Placental Biology Course 2018" --filename "PlacentalBiologyCourse.multiqc_report.html"

firefox PlacentalBiologyCourse.multiqc_report.html
```

FastQC

Version

A quality control tool for high throughput sequence data

0.11.5

Download

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Terminal:

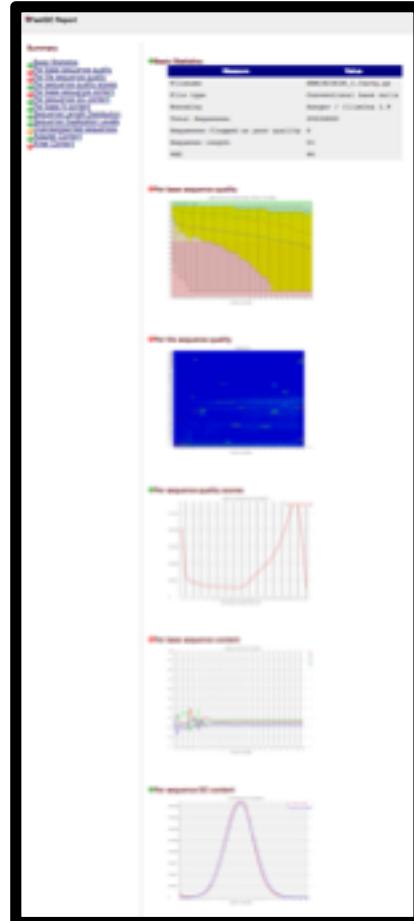
```
Read 1
```

```
Read 2
```

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

```
$ firefox SRR1823638_sub_1_fastqc.html
```

Output:

*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

trim_galore A wrapper tool around Cutadapt to consistently apply quality and adapter trimming to FastQ files
Version 0.4.4
Download http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Terminal:

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

Compress the output fastq files

Read 1

Read 2

--paired

--gzip

-q 20

SRR1823638_sub_1.fastq.gz

SRR1823638_sub_2.fastq.gz

Treat as paired-end

Quality score threshold (PHRED > 20)

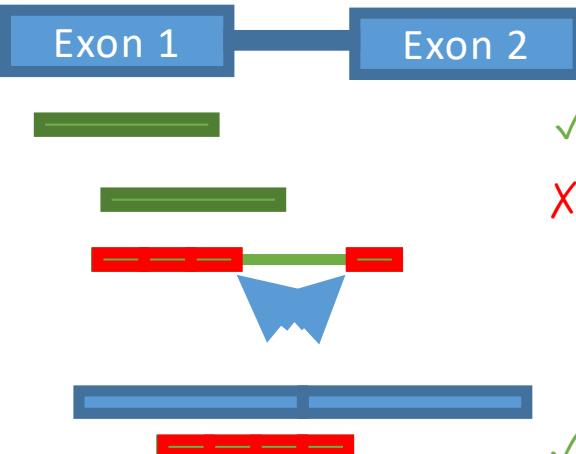
Output:

Trimmed Fastq files

SRR1823638_sub_1_val_1.fq.gz
SRR1823638_sub_2_val_2.fq.gz

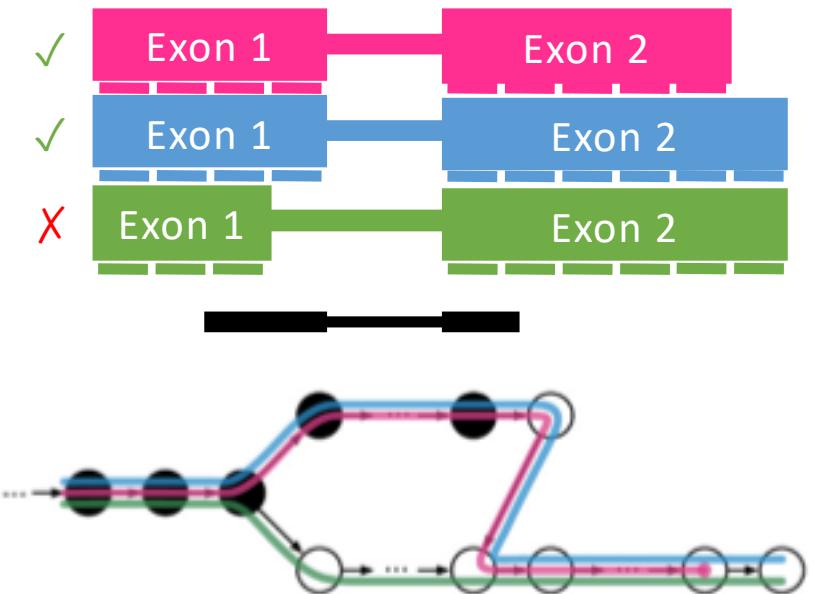
Alignment: STAR Vs Kallisto

STAR: Align to genome



- ✓ Single exon mapping
- ✗ Multi-exon reads
Reads divided into segments
splice site identified
- ✓ Segments aligned and
assembled

Kallisto: Align to transcriptome



| | STAR | Kallisto |
|------------------------------|------------|----------|
| <i>Run time</i> | hours | minutes |
| <i>Hardware requirements</i> | Multi-core | Laptop |
| <i>Novel Splice Sites</i> | yes | no |

kallisto Program for quantifying abundances of transcripts from RNA-Seq data, without the need for alignment
Version 0.43.1
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

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

Normalises the counts for the length of the transcript

Sample normalisation required as performed by DESeq2 and Sleuth

TPM (Transcripts Per Million)

Measurement of the proportion of transcripts in your pool of RNA

| | |
|----------|---|
| MultiQC | Aggregate results from bioinformatics analyses across many samples into a single report |
| Version | 1.1dev |
| Download | http://multiqc.info/ |

Terminal:

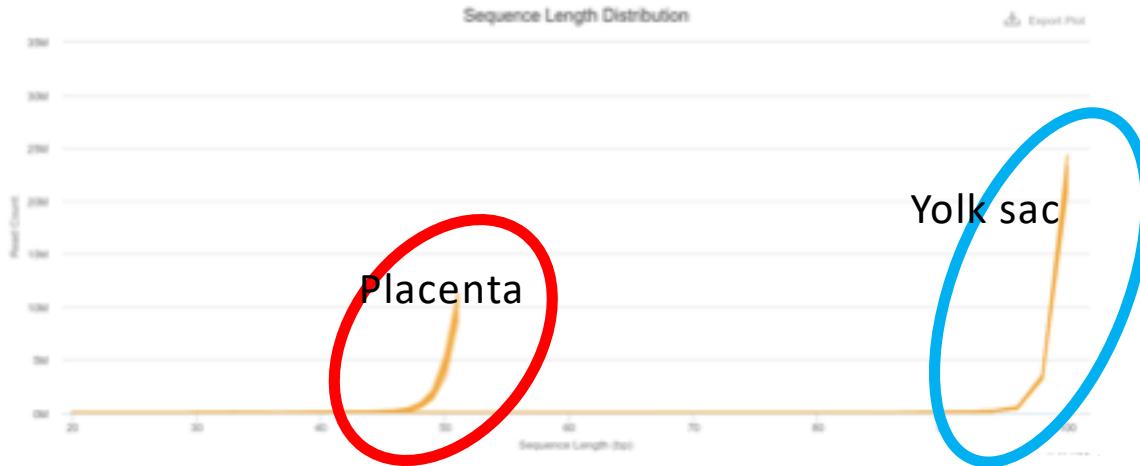
```
$ multiqc -f -i "Placental Biology Course 2018"  
    --filename "PlacentalBiologyCourse.multiqc_report.html" .  
$ firefox PlacentalBiologyCourse.multiqc_report.html
```

Overwrite existing report A title for your report
Output filename ." Is a special Linux symbol which means the current directory

Output:

HTML Report
PlacentalBiologyCourse.multiqc_report.html
PlacentalBiologyCourse.multiqc_report_data

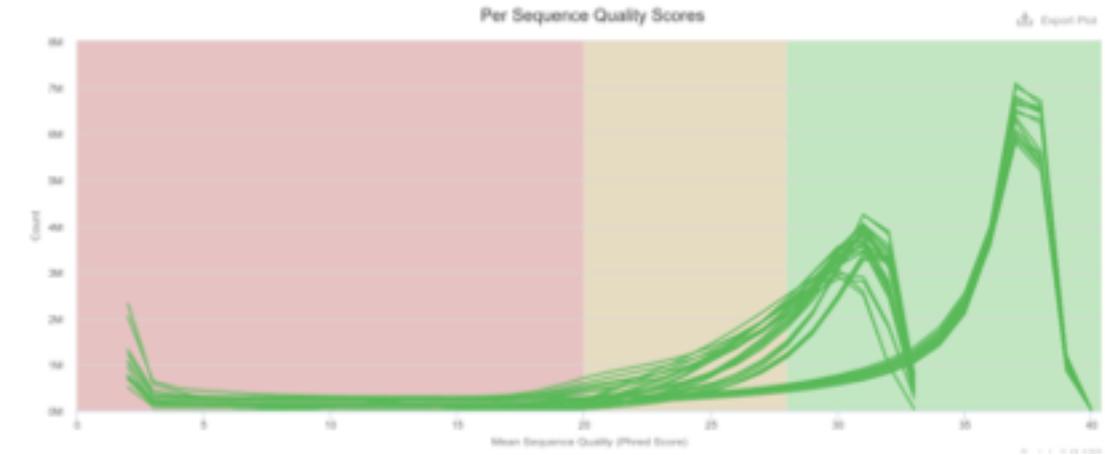
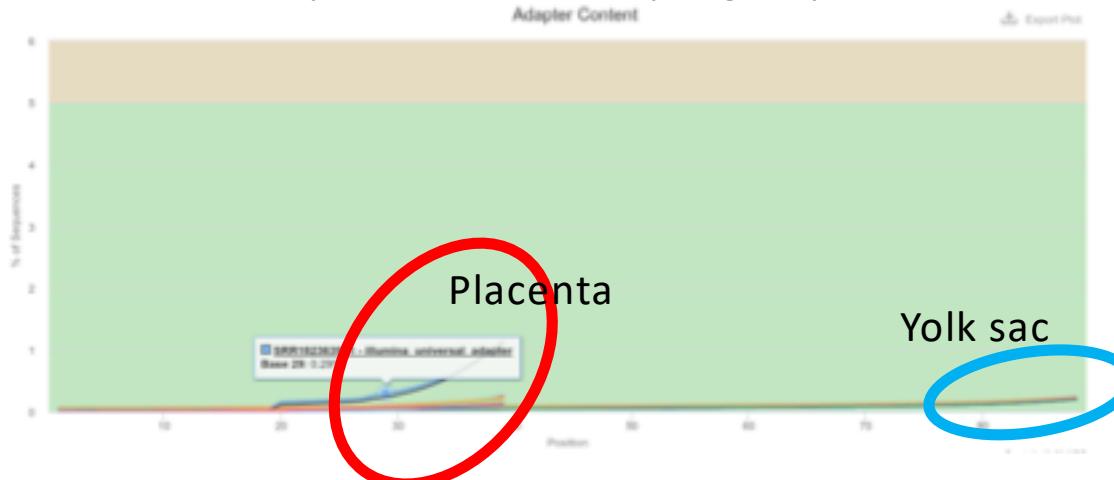
Sample groups have different read lengths



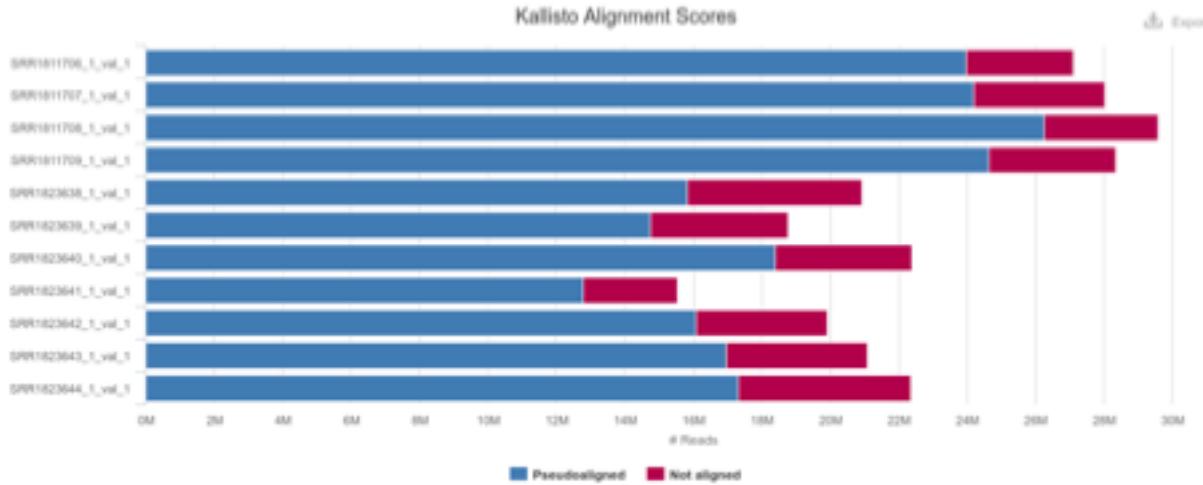
Some Placenta samples have low quality scores



There are adapters in both sample groups



QC Alignments

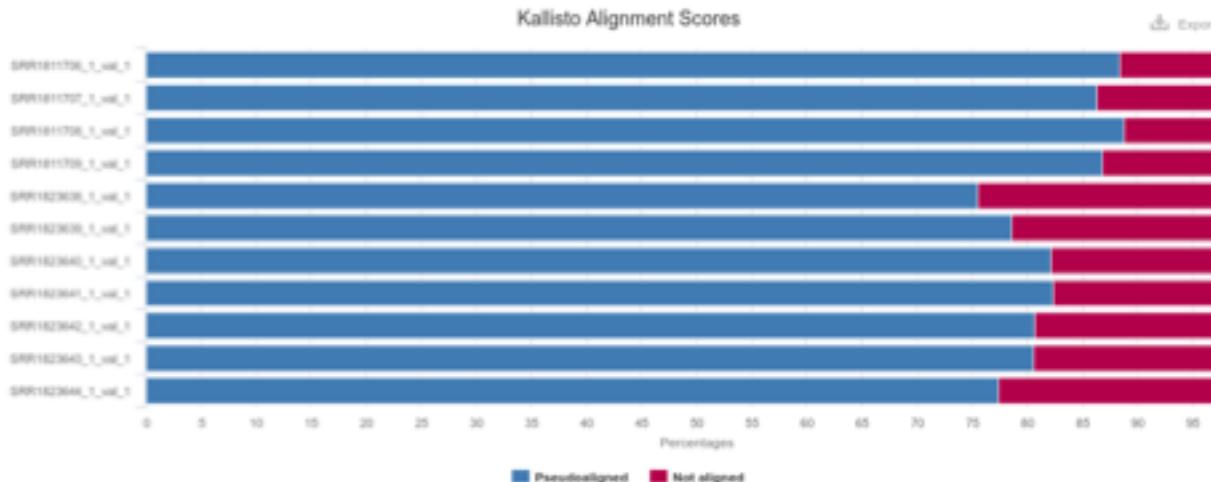


Yolk Sac

Placenta

Why do you never see 100% alignment?

- Incomplete reference genomes / transcriptomes
- Repetitive reads hard to map uniquely
- Sample: Structural Variants Copy Number Variants



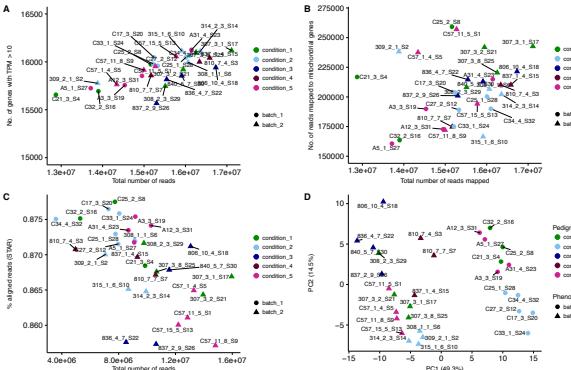
Yolk Sac

Placenta

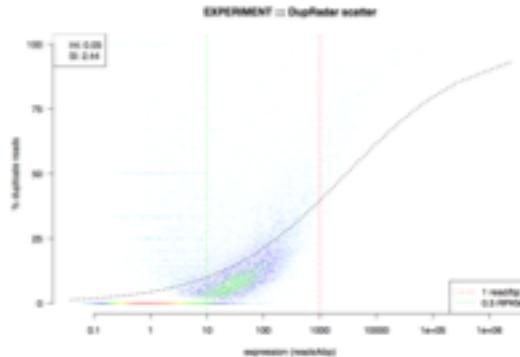
Harsher
trimming,
more reads
removed /
trimmed

Further QC

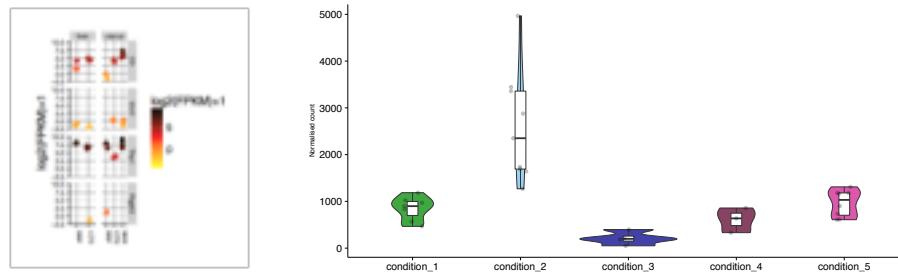
rRNA and mt content



dupRadar



Custom Plots



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



Coffee Break

Part 1. Quality control, alignment and transcript quantification

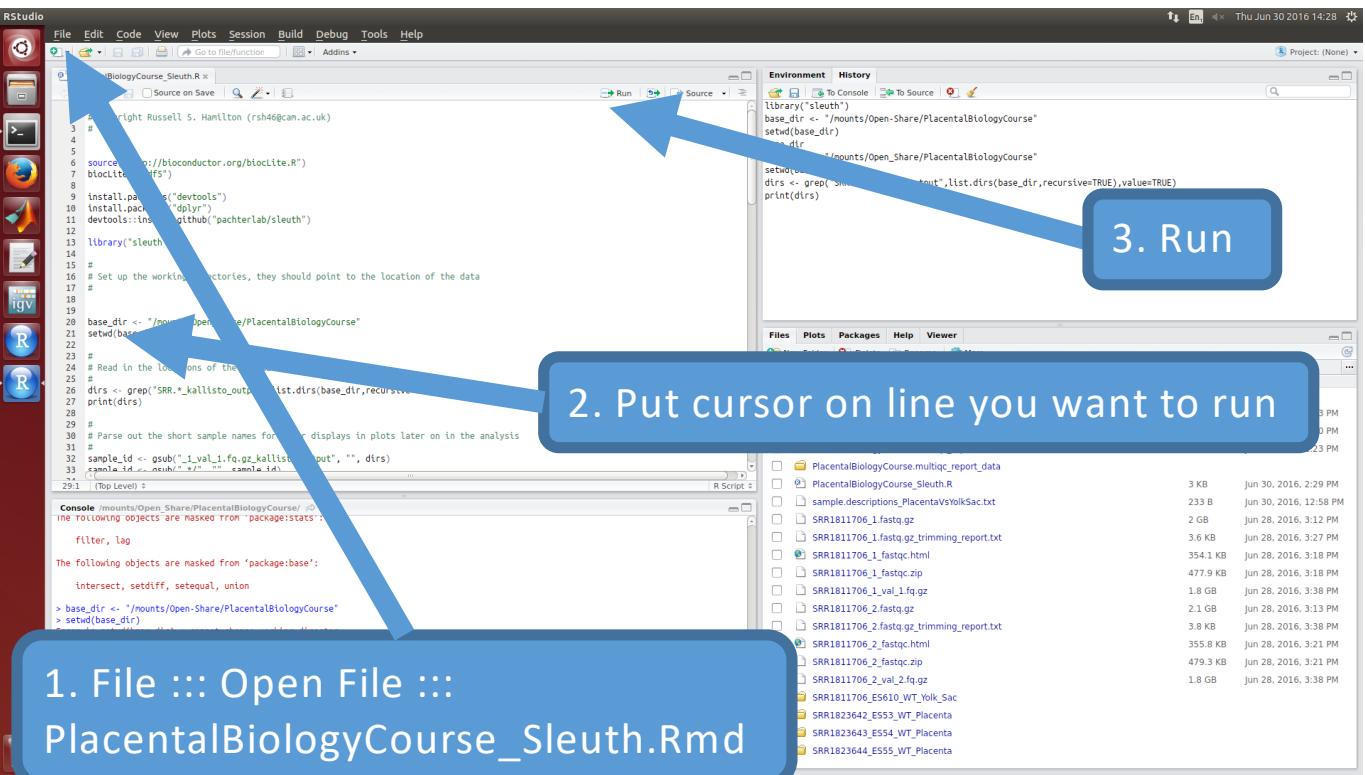
Part 2. Differential Expression Analysis

sleuth

Version

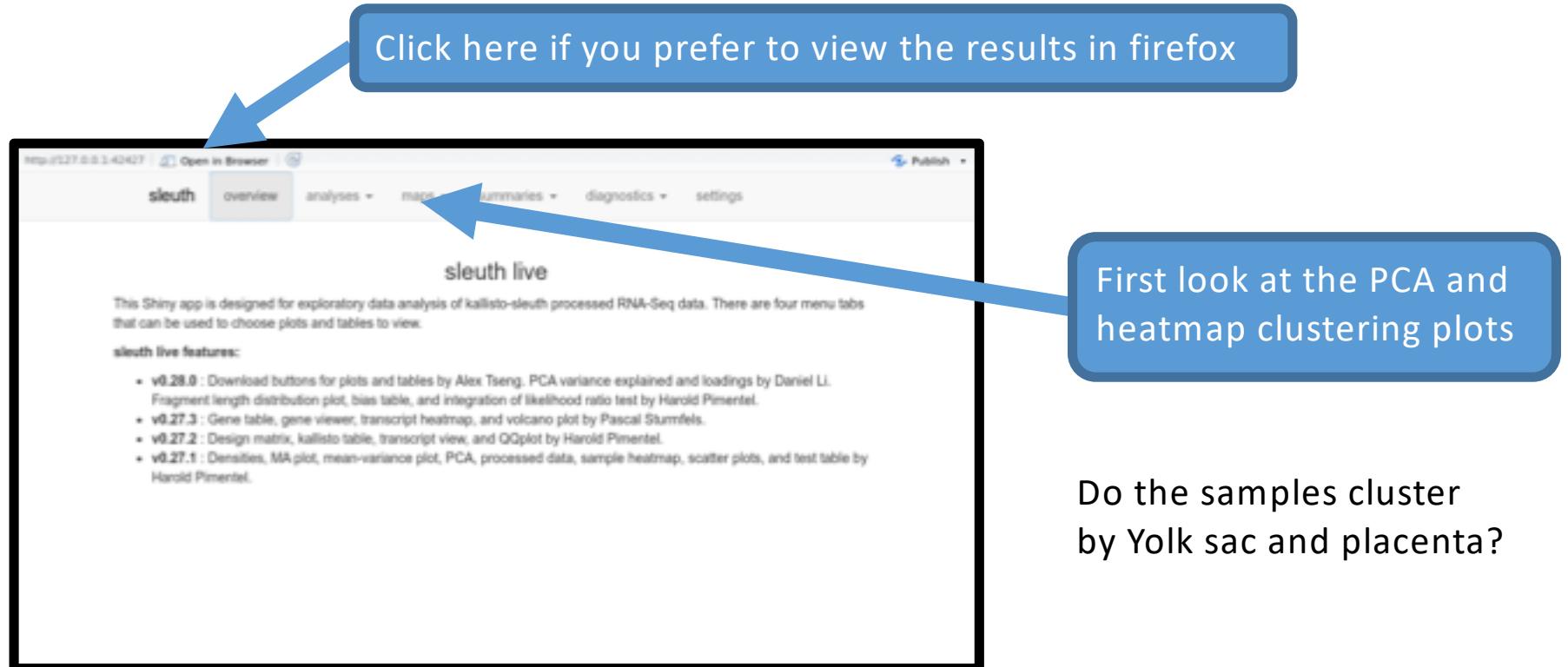
Download

Analysis of RNA-Seq experiments for which transcript abundances have been quantified with kallisto
0.28.1
<http://pachterlab.github.io/sleuth/> R



- A statistical programming language
 - R-Studio, a graphical environment for using R
 - # denotes a comment

Click here if you prefer to view the results in firefox



sleuth live

This Shiny app is designed for exploratory data analysis of kallisto-sleuth processed RNA-Seq data. There are four menu tabs that can be used to choose plots and tables to view.

sleuth live features:

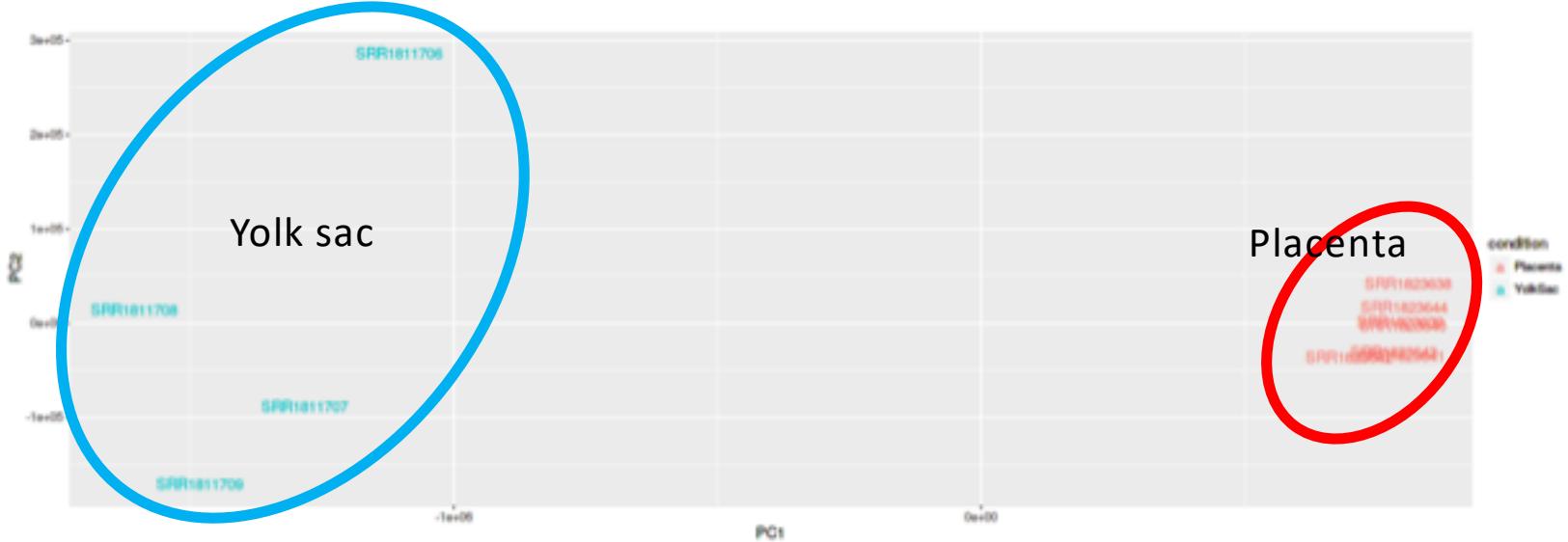
- v0.28.0 : Download buttons for plots and tables by Alex Tseng. PCA variance explained and loadings by Daniel Li. Fragment length distribution plot, bias table, and integration of likelihood ratio test by Harold Pimentel.
- v0.27.3 : Gene table, gene viewer, transcript heatmap, and volcano plot by Pascal Sturmels.
- v0.27.2 : Design matrix, kallisto table, transcript view, and QQplot by Harold Pimentel.
- v0.27.1 : Densities, MA plot, mean-variance plot, PCA, processed data, sample heatmap, scatter plots, and test table by Harold Pimentel.

First look at the PCA and heatmap clustering plots

Do the samples cluster by Yolk sac and placenta?

Sample Clustering

PCA Plot



Heat Map



Volcano Plot

volcano plot

Plot of beta value (regression) versus log of significance. Select a set of transcripts to explore their variance across samples.



Differentially Expressed Genes

Paste an external gene
shortcode here

gene view

(Boxplots of abundances of transcript mapping to a given gene, and their technical variation. This step can take a while, especially with many plots.)

gene:

ENSGALUS000000022150.16

color by:

condition

units:

tpm

genes from:

ens_gene

of plots (max 15):

3

view

ENSMUST00000160134.7

Yolk sac

Placenta

sample

ENSMUST00000080880.11

Yolk sac

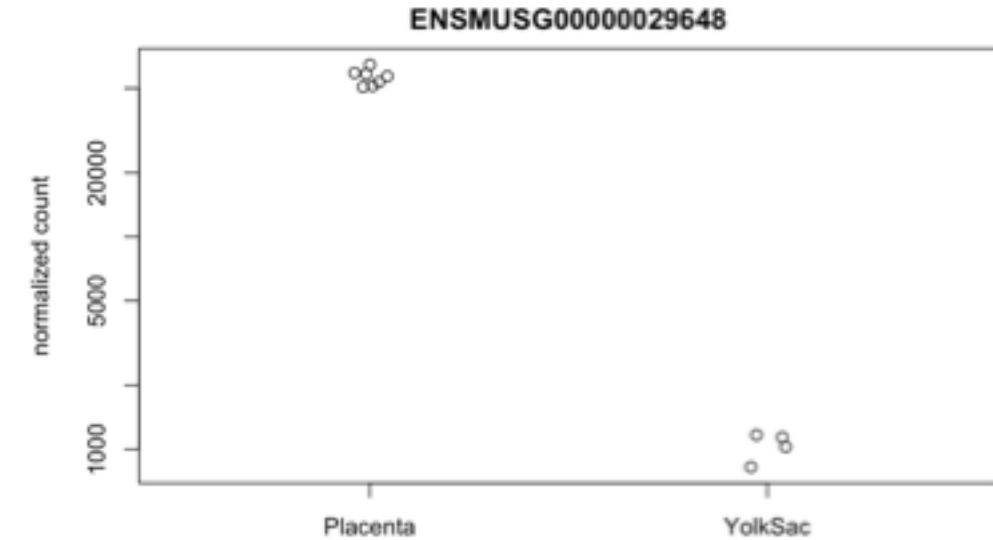
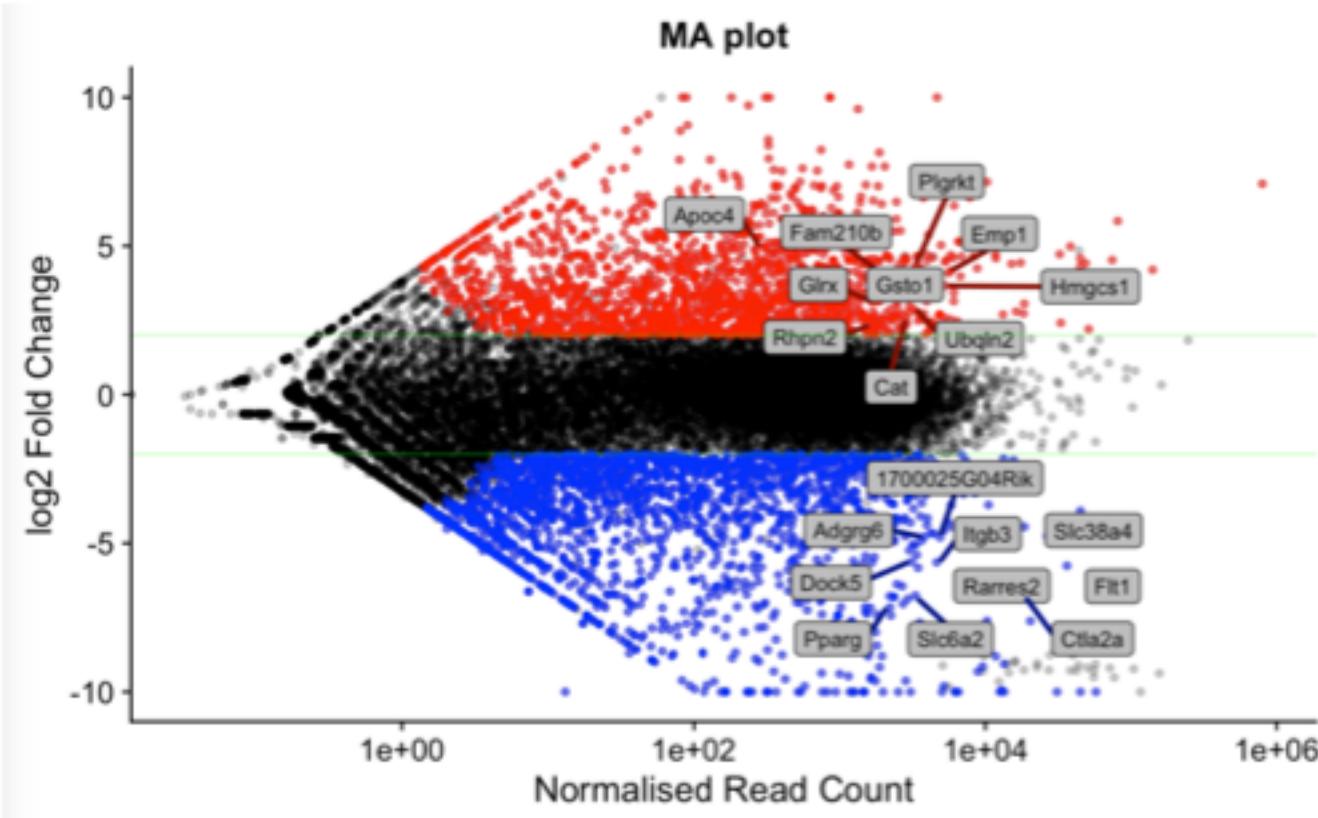
Placenta

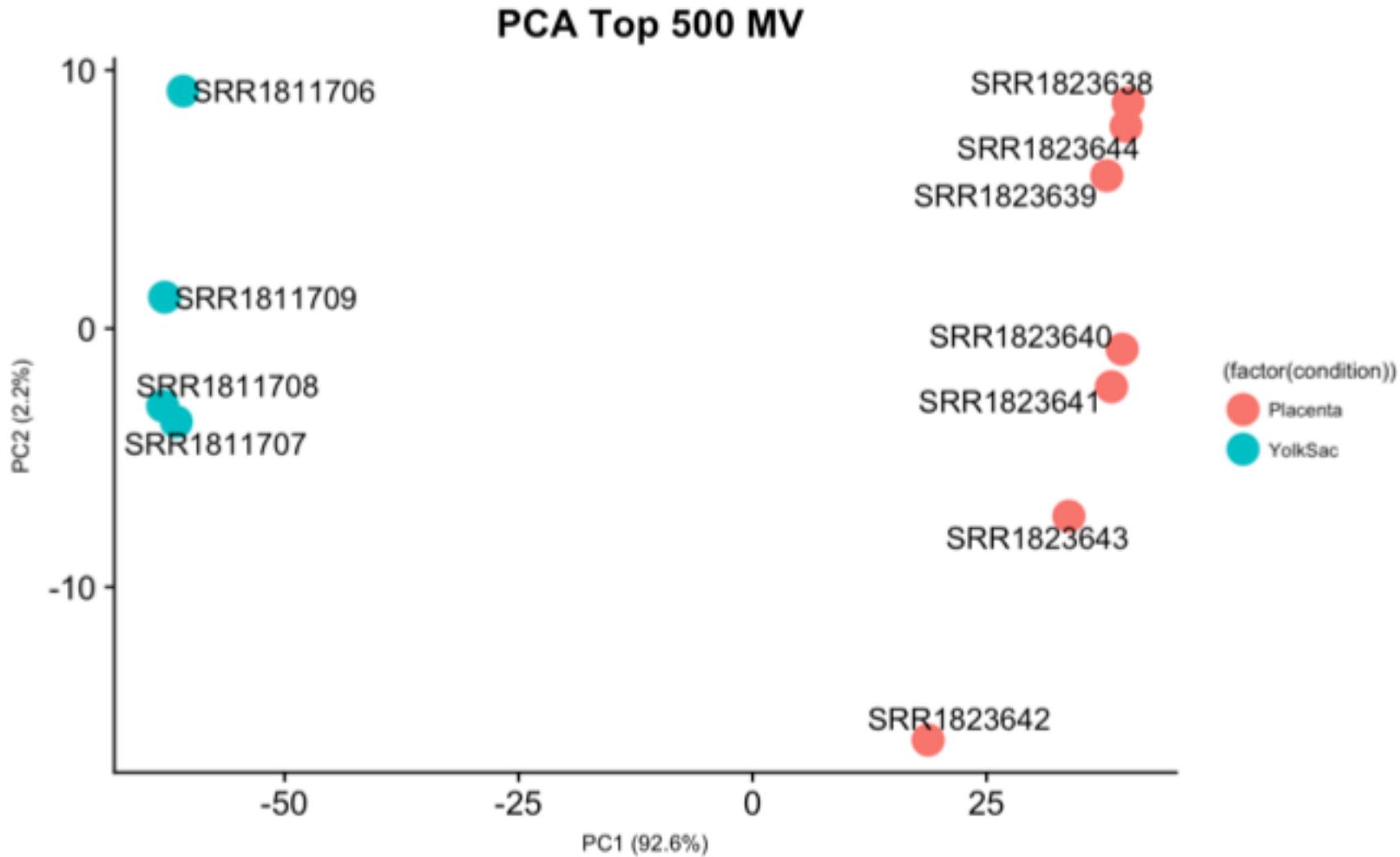
sample

Select TPM
Transcripts Per Million

Select external_
gene_ shortcode

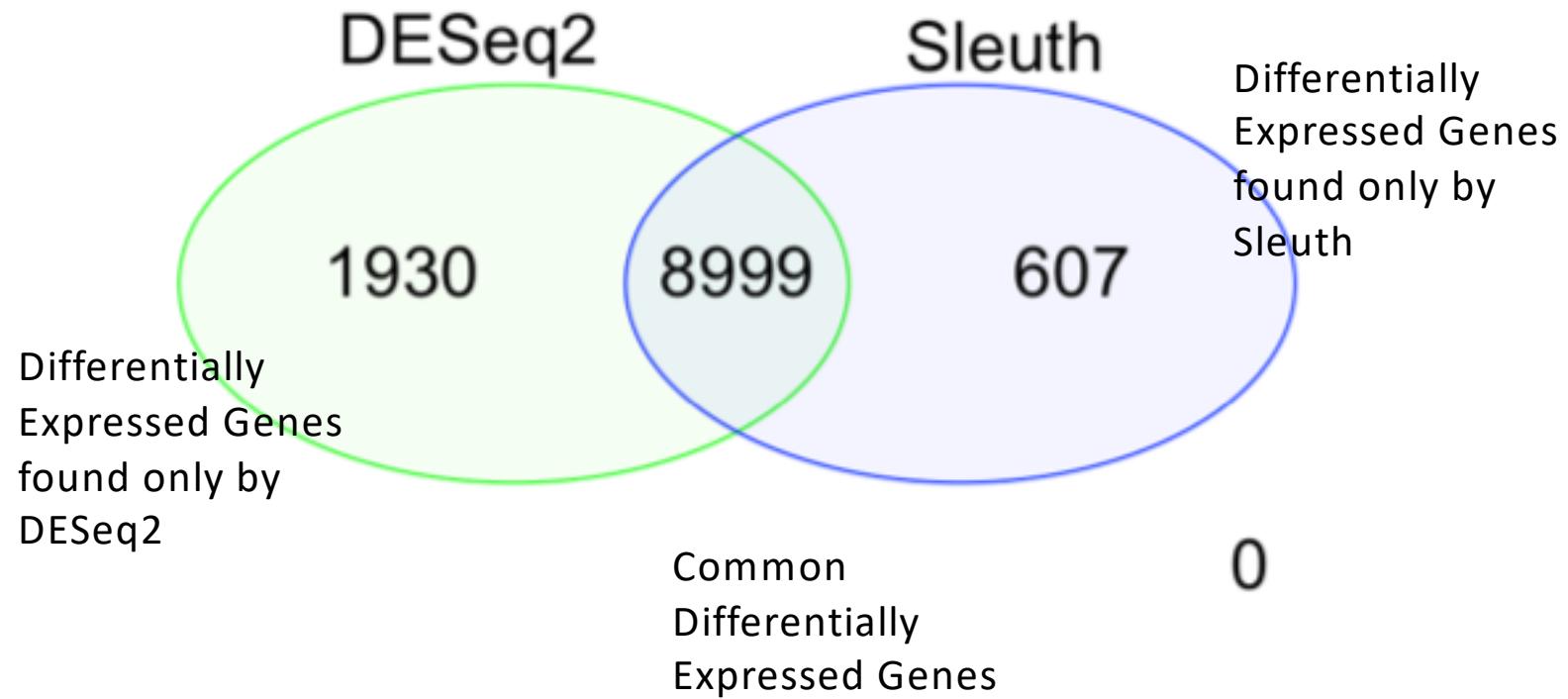
From global picture to individual genes





Similar to the Sleuth version, but you can customise number of genes to use, colour, labels etc much more easily using the DESeq2/ggplot custom version

Differentially Expressed Genes



- DESeq2 is considered to perform better at a gene level
- Sleuth performs better on a transcript level
 - Better for identifying alternatively spliced isoform usage

Looking at the Alignments

File

Load From File...

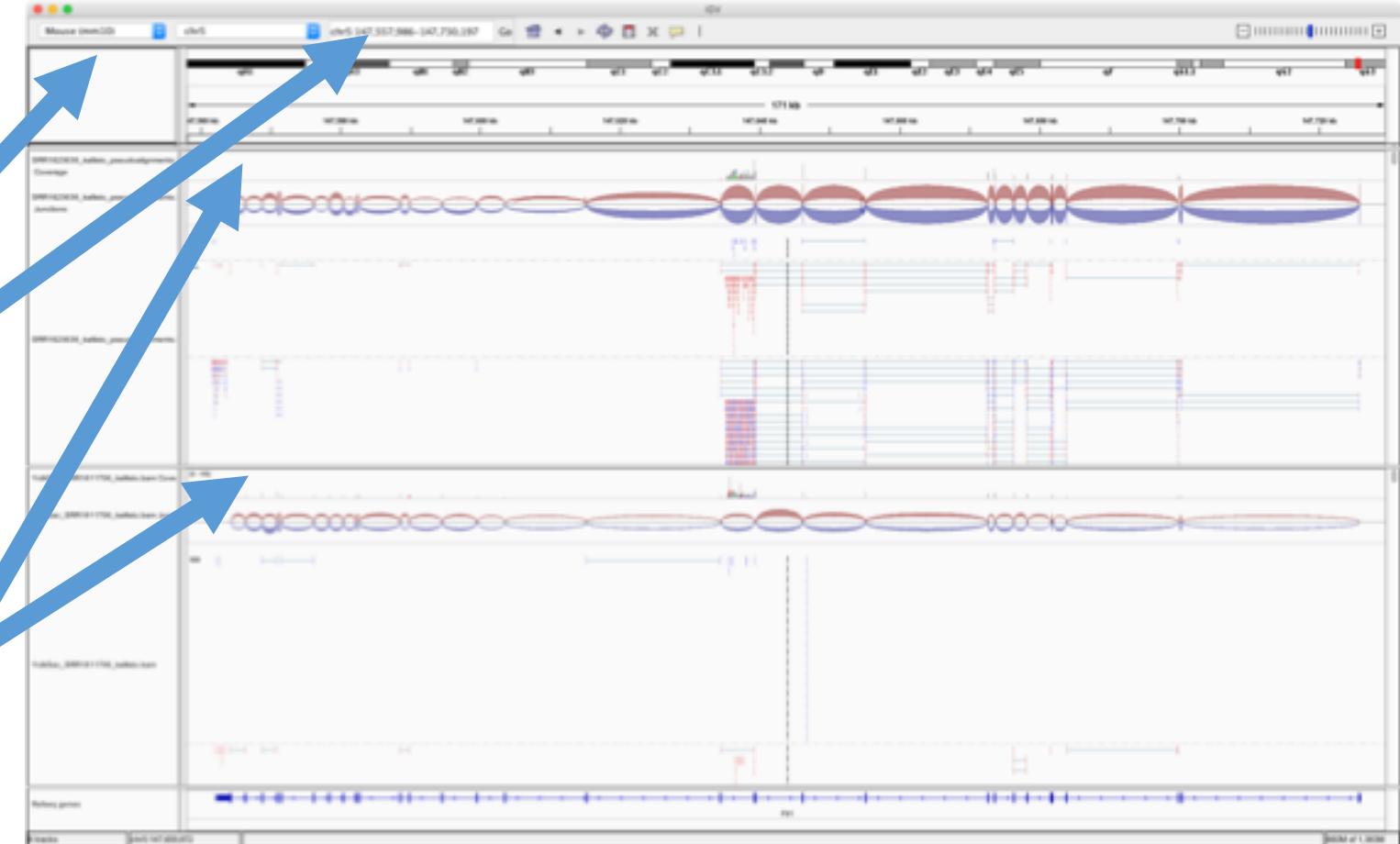
IGV_Alignments/

Placenta_SRR1823638_kallisto.bam
~~Placenta_SRR1823638_kallisto.bam.bai~~
YolkSac_SRR1811706_kallisto.bam
~~YolkSac_SRR1811706_kallisto.bam.bai~~

Select: Mouse(mm10)

Search for Flt1

Can you see the difference in coverage?



http://www.ensembl.org/Mus_musculus/Location/Genome

Enter gene to search here
e.g. Trf



The Ensembl genome browser for *Mus musculus* (GRCm38) - [Gene ID: 114](#) - [Search](#)

Gene-based analysis

- Summary
- Update variants
- Transcript comparison
- Protein comparison
- Gene details
- Sequence
- Secondary Structure
- Expression summaries
- Evolution
- Diseases
 - GO Biological process
 - GO Cellular component
 - GO Molecular function
- Comparative genomics
 - Genomic alignments
 - Conservation scores
 - Gene phylogenetic tree
 - Orthologues
 - Proteins
 - Protein families
- Genetic variation
 - Variant lists
 - Variant tracks
 - SNP variants
- External data
 - Gene expression
 - Proteins
 - ChIP-seq
- Configure file input
- Add your data
- Export data
- Show this page
- Bookmark this page

Gene: Trf [ensembl.gene](#)

Description: transforming factor beta 1, isoform 1A

Synonyms: Trf, Ctnn1, 40300001, Igpc, Tf, Tf-1, Tf-1A

Location: Chromosome 2, 125,000,000-125,200,000 (mouse.m38), GRCm38.Chr2.125.000.000-125.200.000

About this gene

This gene has 11 transcripts (splice variants), 15 orthologues, 12 aliases. It is a member of [Ensembl protein family](#) and is associated with [Lphenotrans](#).

Transcripts

[View transcript details](#)

| Genebody columns (1 hidden) | Transcript ID | Str. | Protein | Biotype | Start | End | Strand | RefSeq | Proteins |
|-----------------------------|--------------------|------|---------|----------------|-----------|-----------|--------|------------------------------|----------|
| | ENSMUST00000120411 | 1 | | Protein coding | 125000000 | 125023474 | - | NC_000000000.12 NP_001000711 | |
| | ENSMUST00000120409 | 2 | | Protein coding | 125000000 | 125023474 | - | NC_000000000.12 NP_001000711 | |
| | ENSMUST00000120408 | 3 | | Protein coding | 125000000 | 125023474 | - | NC_000000000.12 NP_001000711 | |
| | ENSMUST00000120407 | 4 | | Protein coding | 125000000 | 125023474 | - | NC_000000000.12 NP_001000711 | |
| | ENSMUST00000120406 | 5 | | Protein coding | 125000000 | 125023474 | - | NC_000000000.12 NP_001000711 | |
| | ENSMUST00000120405 | 6 | | Protein coding | 125000000 | 125023474 | - | NC_000000000.12 NP_001000711 | |
| | ENSMUST00000120404 | 7 | | Protein coding | 125000000 | 125023474 | - | NC_000000000.12 NP_001000711 | |
| | ENSMUST00000120403 | 8 | | Protein coding | 125000000 | 125023474 | - | NC_000000000.12 NP_001000711 | |
| | ENSMUST00000120402 | 9 | | Protein coding | 125000000 | 125023474 | - | NC_000000000.12 NP_001000711 | |
| | ENSMUST00000120401 | 10 | | Protein coding | 125000000 | 125023474 | - | NC_000000000.12 NP_001000711 | |
| | ENSMUST00000120400 | 11 | | Protein coding | 125000000 | 125023474 | - | NC_000000000.12 NP_001000711 | |

Summary

Name: [Trf](#) (SYN0001)

ENSG00000000000

orthologs

RefSeq

Ensembl version

Gene type

Annotation method

Alternative genes

Annotation attributes

overlapping known databases

Go to Region in Detail for more tracks and navigation options (e.g. zooming)

Gene track: [Trf](#) (SYN0001)

Chromosome 2: 125,000,000 - 125,200,000

Genomic coordinates: 125,000,000 - 125,200,000

Gene Legend:

- Protein Coding
- Known protein coding
- Unspliced transcript

What is the function of Trf?

DAVID Gene Ontology Analysis

<https://david.ncifcrf.gov/>

Start Analysis



The screenshot shows the DAVID Bioinformatics Resources 6.8 homepage. It features a top navigation bar with links like Home, Start Analysis, About DAVID Tools, Technical Center, Downloads & APIs, Terms of Service, Why DAVID?, and About Us. Below this is a banner with a green DAVID logo and the text "Welcome to DAVID 6.8 with updated Knowledgebase (more info) *** If you are looking for DAVID 6.7, please visit our development site. ***". A main content area titled "Welcome to DAVID 6.8" includes a search bar, a "What's Important in DAVID?" section with a bulleted list, a "Statistics of DAVID" section with a bar chart, and a "List of DAVID Features" section with a bulleted list. At the bottom are three "Screen Shot" thumbnails.

Paste List of ensEMBL gene identifiers



The screenshot shows the DAVID Analysis Wizard interface. It has a top navigation bar with links like Home, Start Analysis, About DAVID Tools, Technical Center, Downloads & APIs, Terms of Service, Why DAVID?, and About Us. Below this is a banner with a green DAVID logo and the text "Welcome to DAVID 6.8 with updated Knowledgebase (more info) *** If you are looking for DAVID 6.7, please visit our development site. ***". A central "Analysis Wizard" section contains three steps: Step 1 (Submit your gene list through left panel), Step 2 (Choose from a file or Multi-List File), and Step 3 (Select Identifier). Step 1 is currently active, showing a text input field with "1487_A_at" and a "Save" button. Step 2 shows a "Choose File" button and a "Multi-List File" radio button. Step 3 shows a "Select Identifier" dropdown menu with "AFFYMETRIX_SPARE_V7_ID" selected. A large blue arrow points from the "Paste List of ensEMBL gene identifiers" callout to the "Save" button in Step 1.

Select gene list
and GO!



Reproducible Bioinformatics

Versioning

If you write code or scripts use a versioning system (a bit like track changes in Word)
Make it publicly available so people can comment and submit bug reports
e.g. <http://www.github.com>

Pipelines

Track program version numbers, consistent processing and reporting
Avoid manual input of data or settings
e.g. <http://clusterflow.io> or SnakeMake

Data Repositories

Upload your published data to GEO, ENA, SRA etc



Dr Russell S. Hamilton

Email: rsh46@cam.ac.uk

Dr Xiaohui Zhao

Email: xz289@cam.ac.uk

Dr Malwina Prater

Email: mn367@cam.ac.uk



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.