



Introduction to RNA-Seq

Thomas D. Otto



Overview

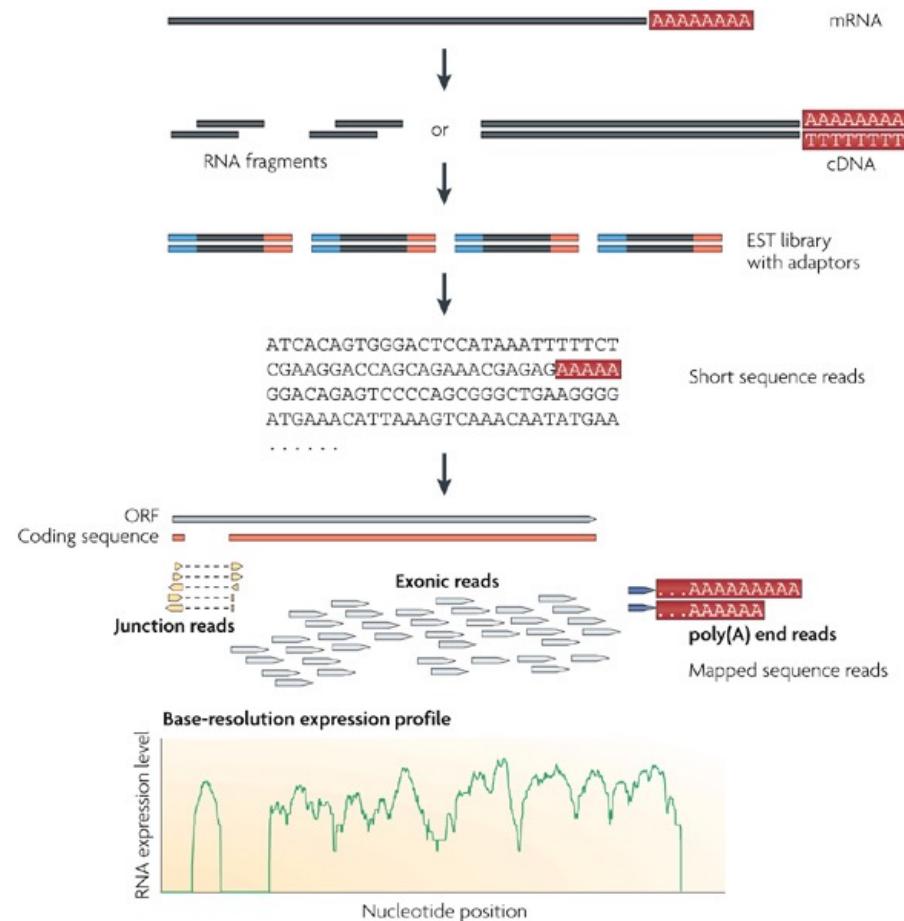
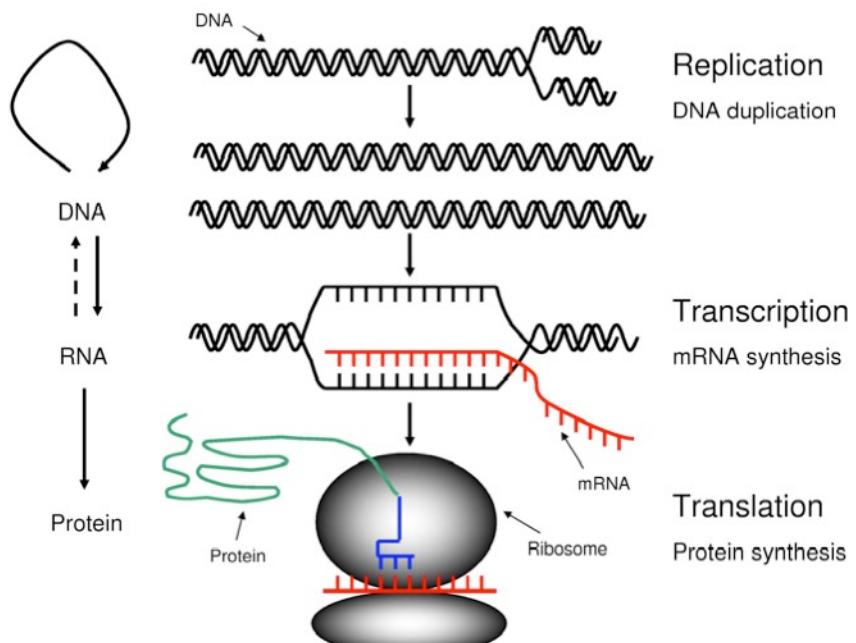
- Introduce RNA-Seq
 - Differential expression
 - Cluster analysis
-
- Go enrichment
 - Functional analysis

Why do we want to sequence the transcriptome?

Why do we want to sequence the transcriptome?

- Capture full unbiased view of full repertoire of transcripts
- Explore alternative splicing, differential expression, complex patterns of expression and regulation
- Functional studies for conditions such as stress and drug resistance
- Pathogen-Host interaction
- Parasite expression atlas
- Possible to work without a genome reference

Transcriptome sequencing



WTAC 2022
What does a sequencing read represents?

considerations before you start

- Do you have enough biological replicates?
- Do you have enough RNA?
- Is the RNA what you want?
- Are you using an appropriate selection method?
- Are you interested in strand differentiation?
- Do you have appropriate controls?

RNA extraction

Do you have enough RNA?

Is the RNA what you want

- Host vs pathogen?

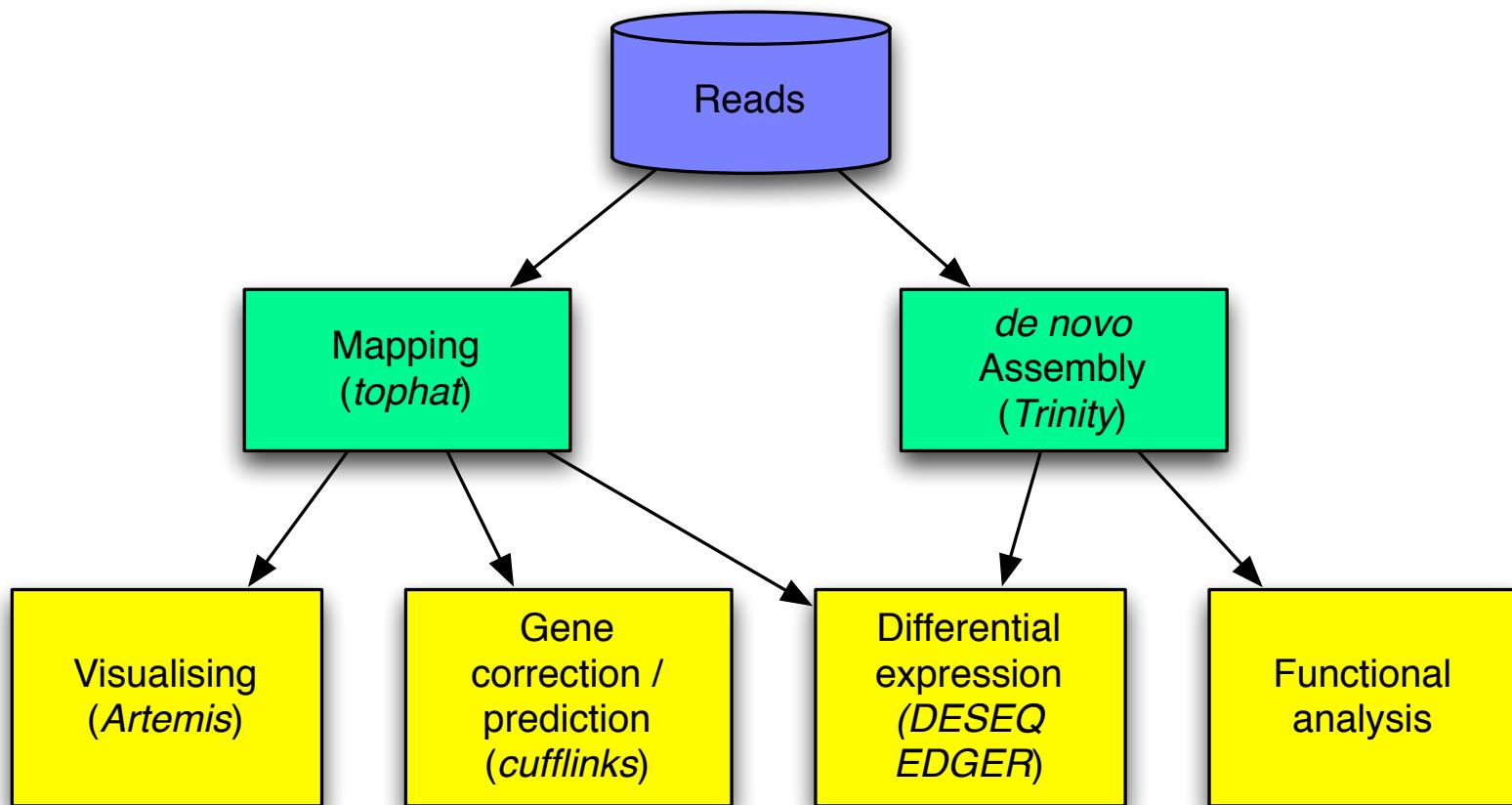
- Protein coding vs non-coding?

Are you using an appropriate selection method?

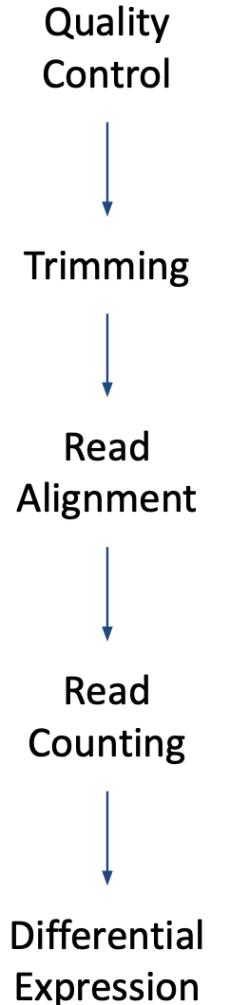
Are you interested in strand differentiation?

What stages or tissues do you want to sample?

Pipeline



Differential expression analysis



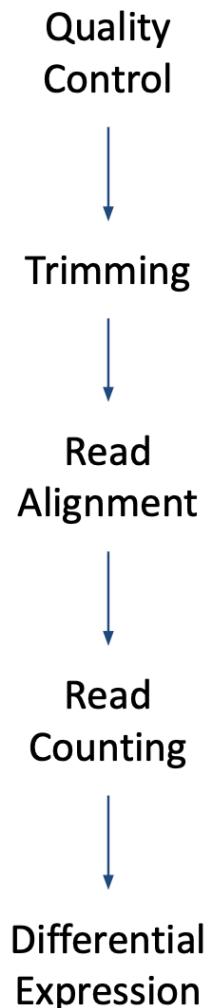
How to map reads

reference: GGGTTAGCGATGGAGA

read1: GA

read2: TAGCG

read3: GGCG



How to map reads

Read

TAGCG

|||||

GGGTTAGCGATGGAGA

Reference

Read

GG-----CG

|| ||

GGGTTAGCGATGGAGA

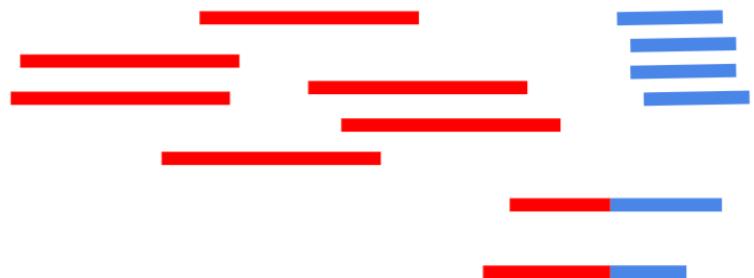
Reference

Splicing

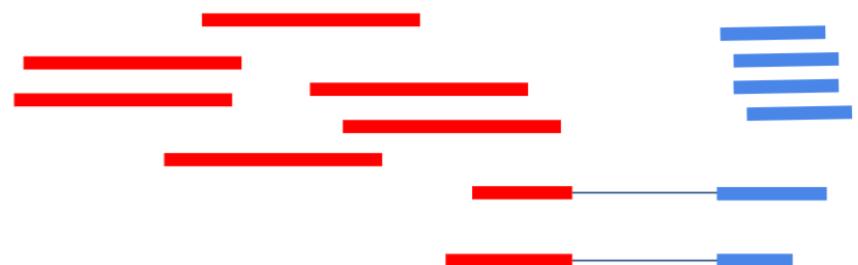
Mapping RNA-Seq



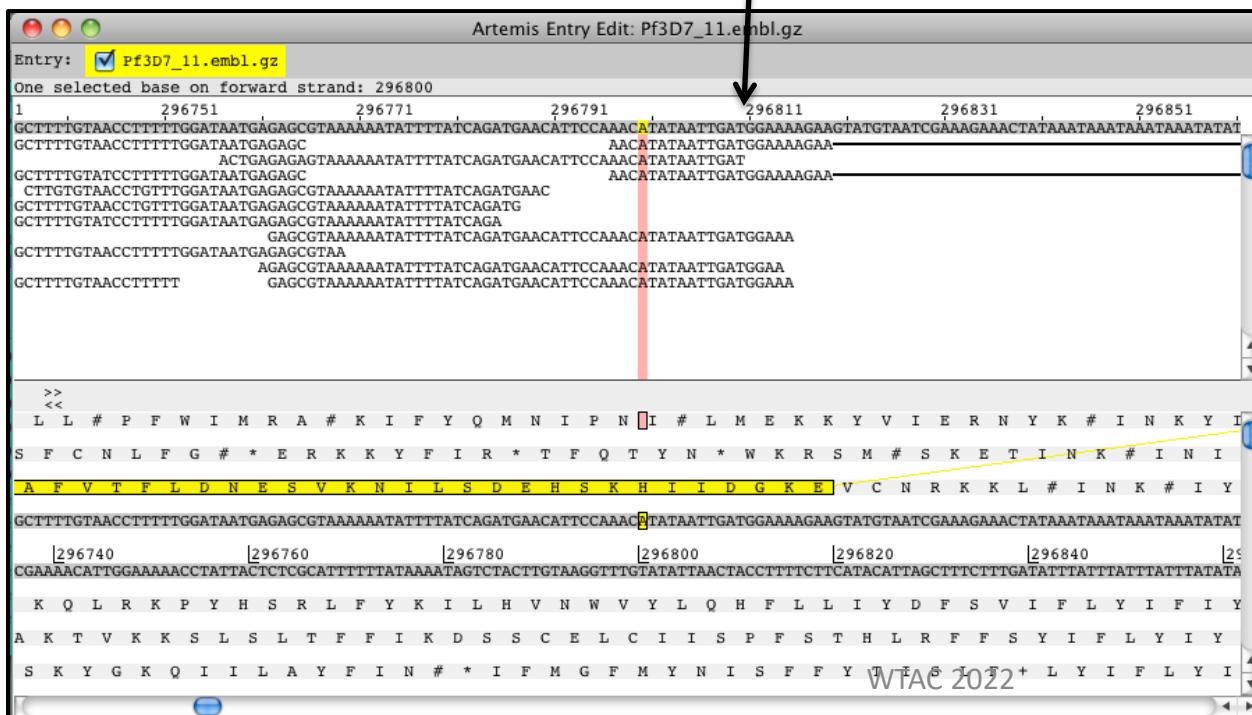
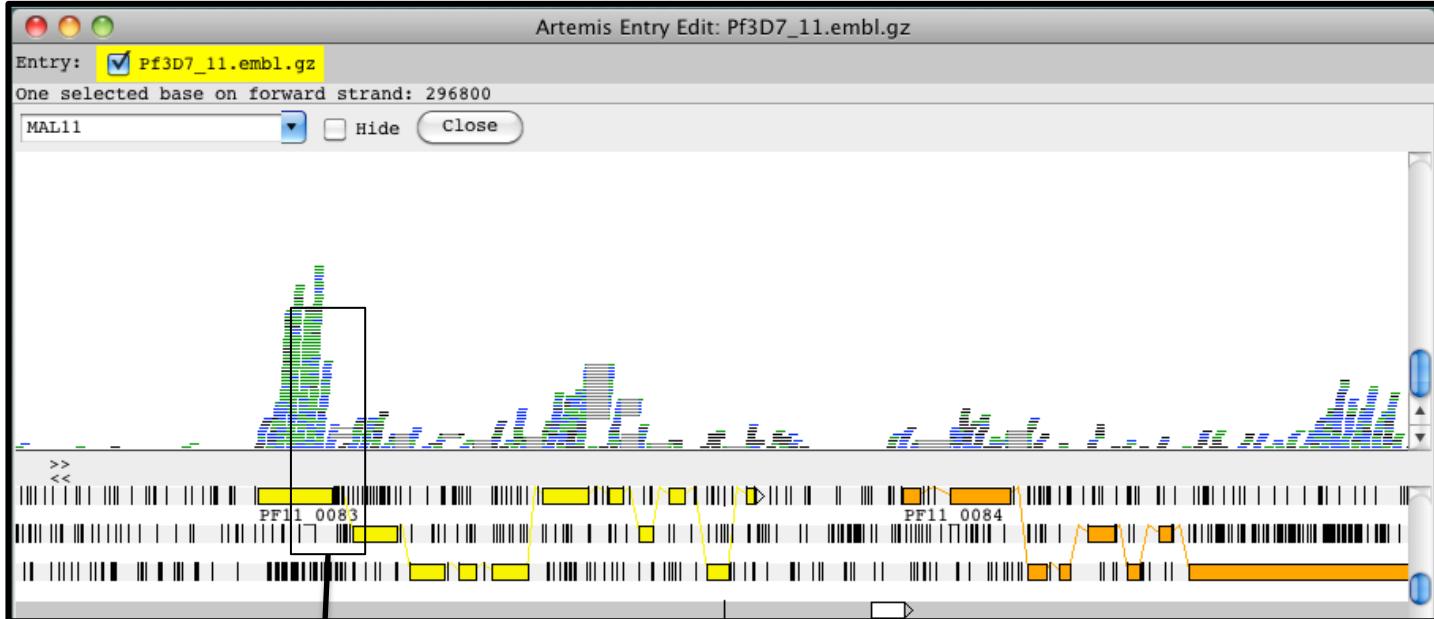
Spliced mRNA



Reference genome



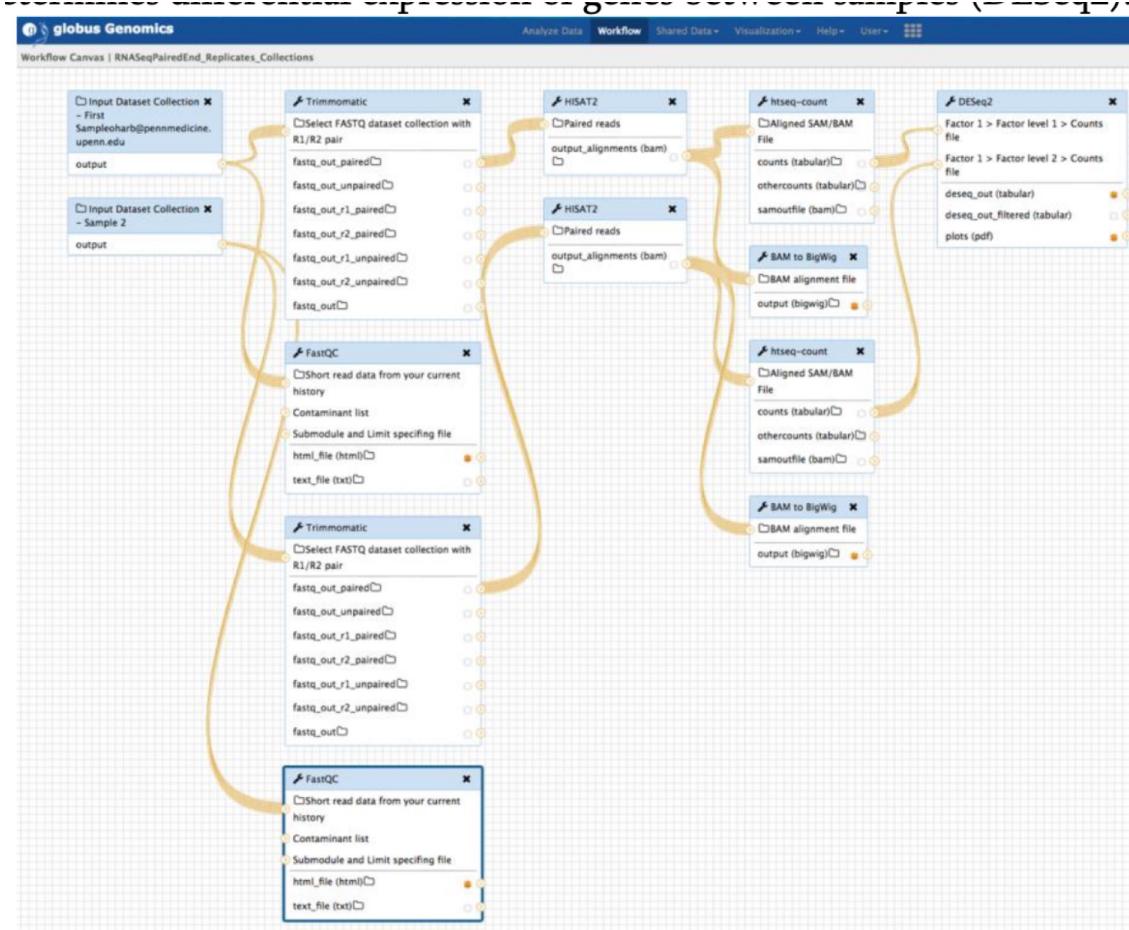
Mapped reads in Artemis



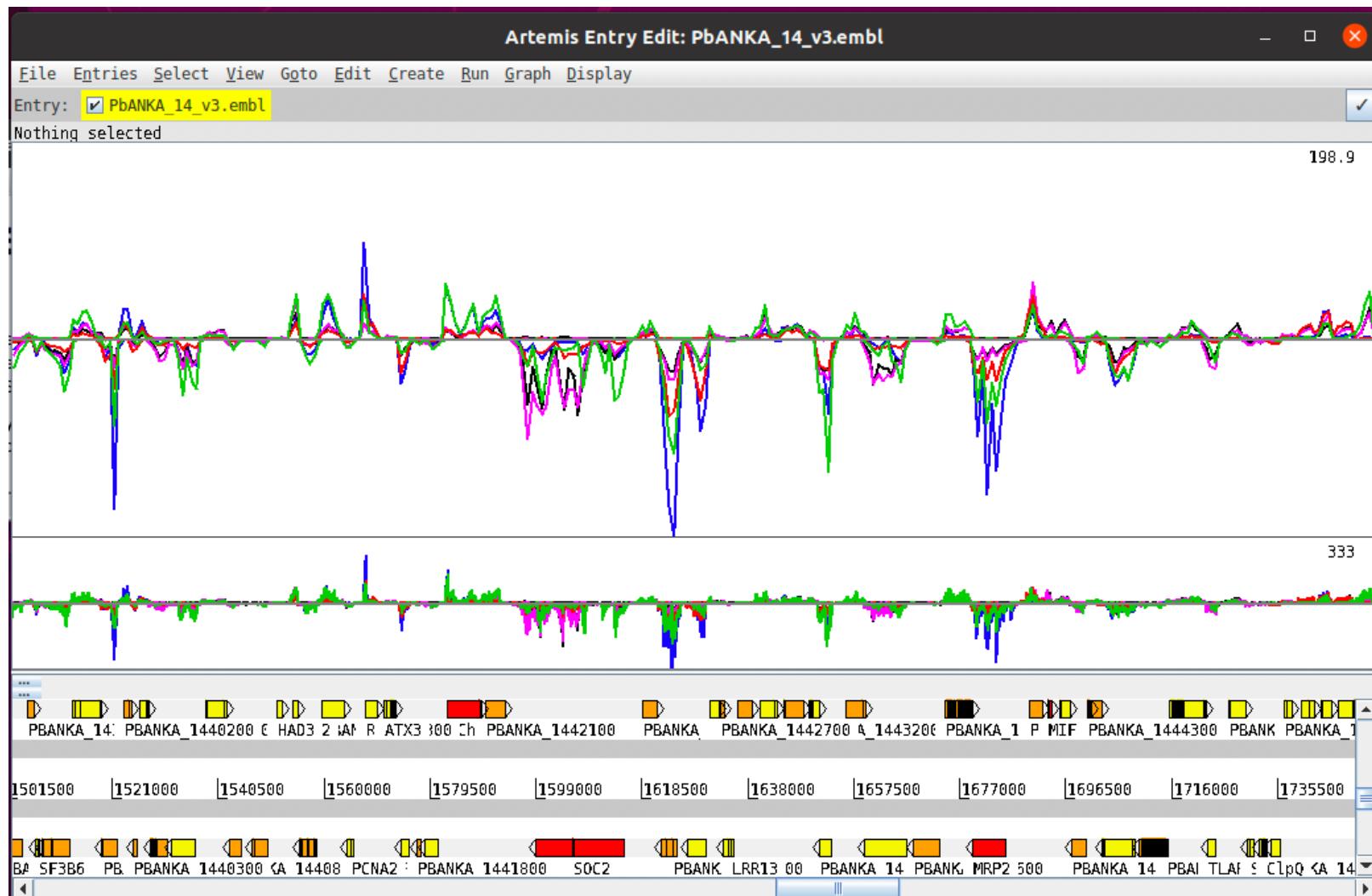
What mappers are spliced aware?

- HiSat2
- STAR
- GMAP
- Pseudo alignment: Kallisto / Salmon

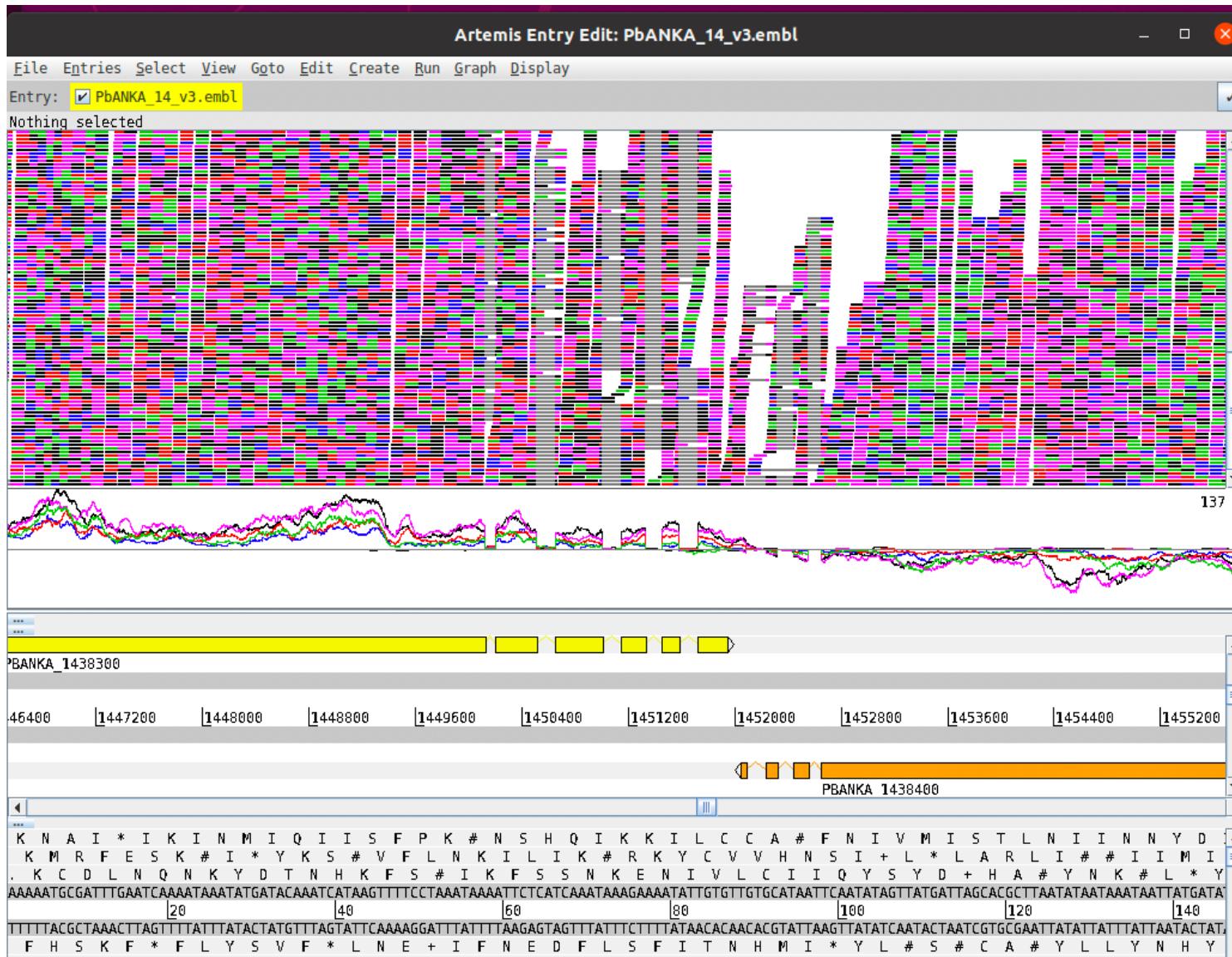
Galaxy workflow



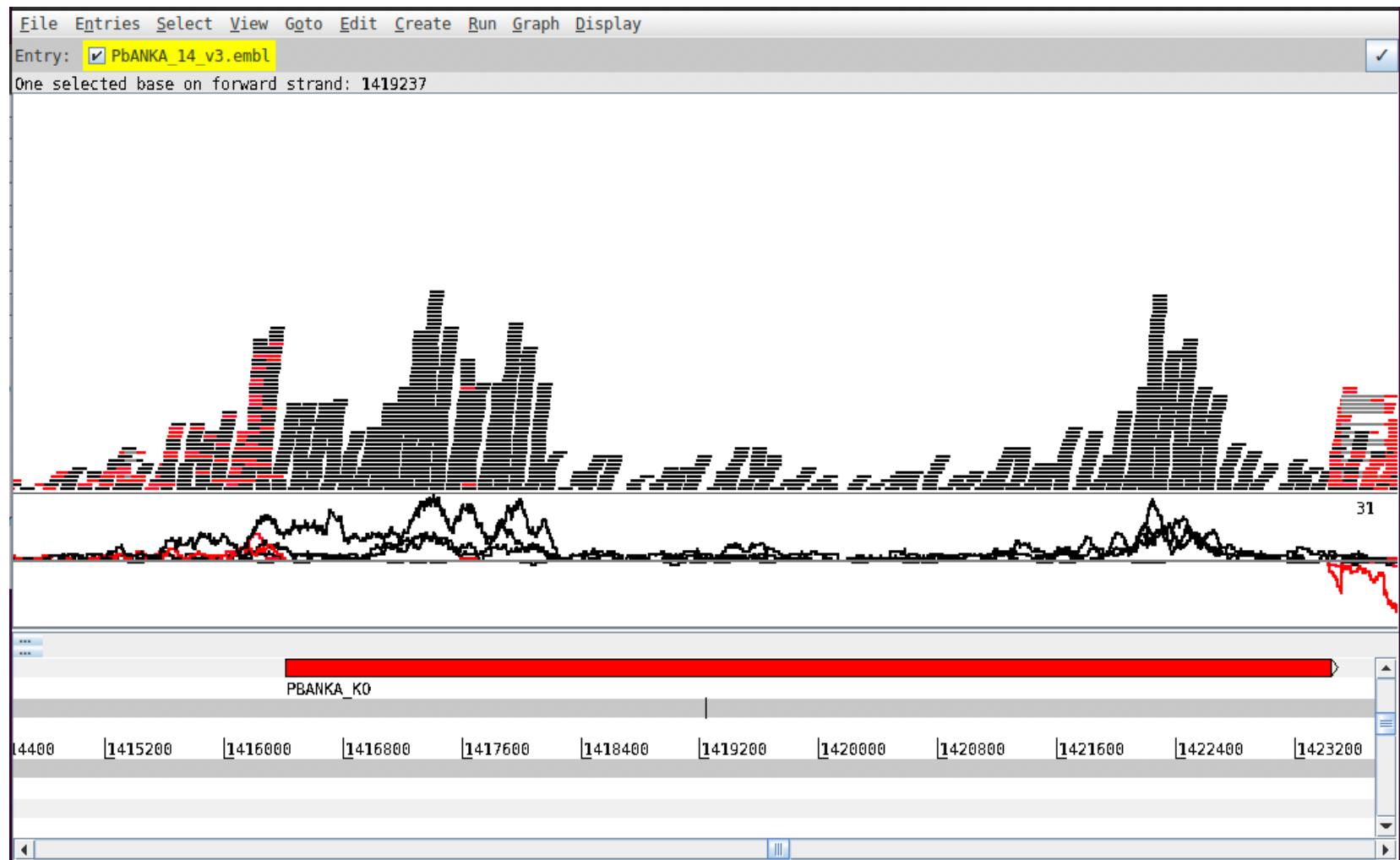
How does it look like in Artemis?



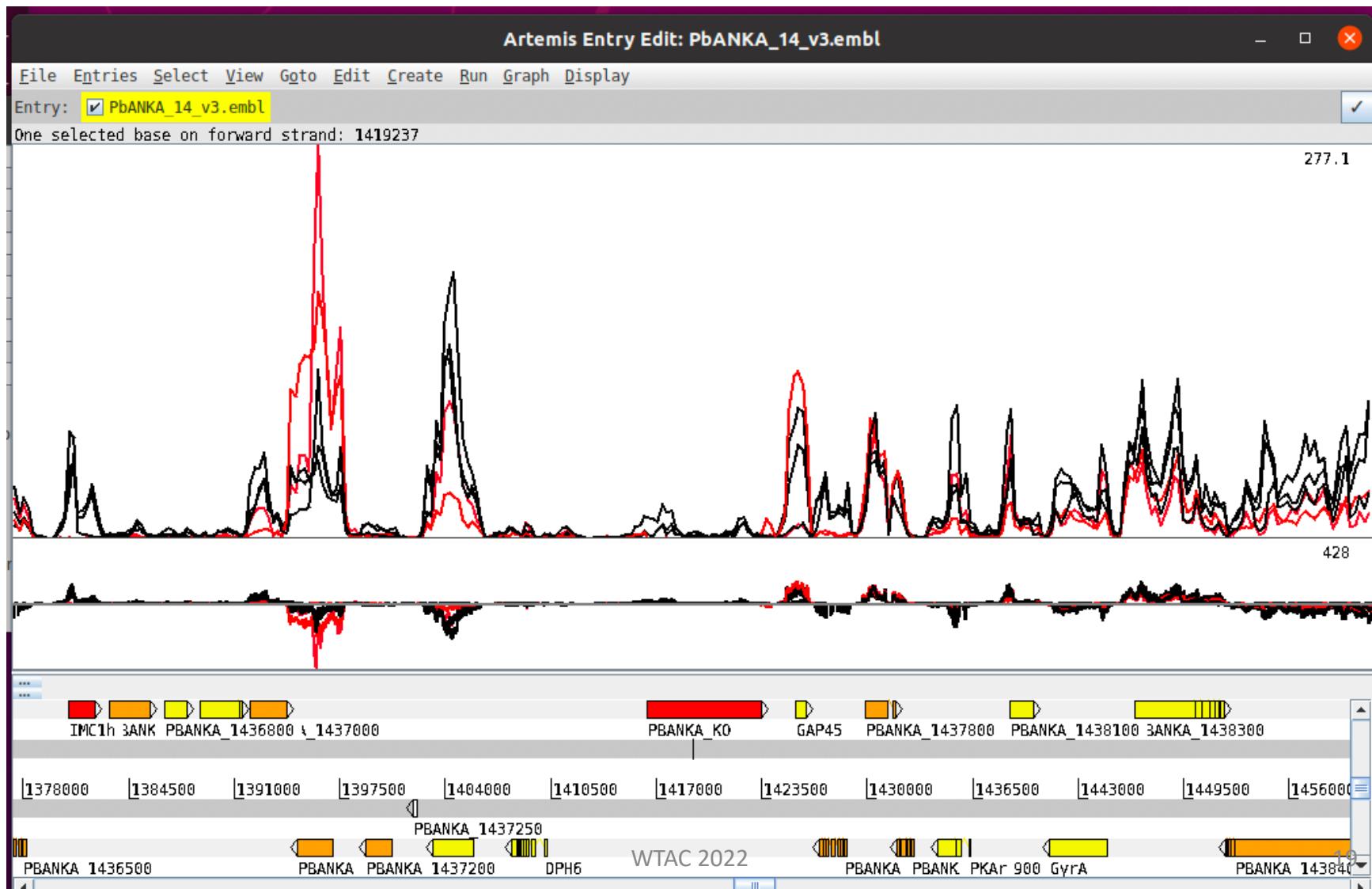
What can we learn?



Application – What is different after KO



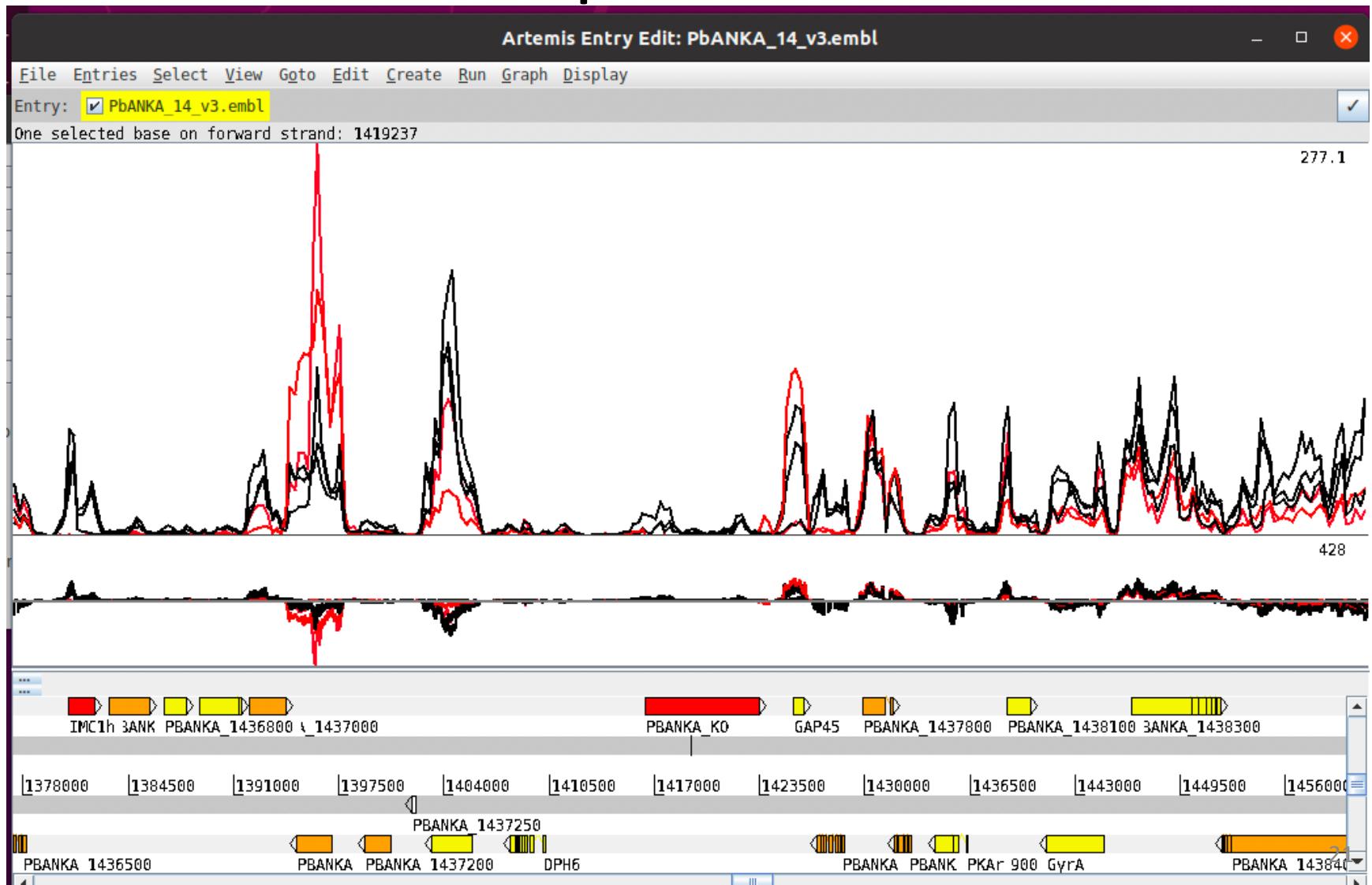
Application – is it enough to look in Artemis?

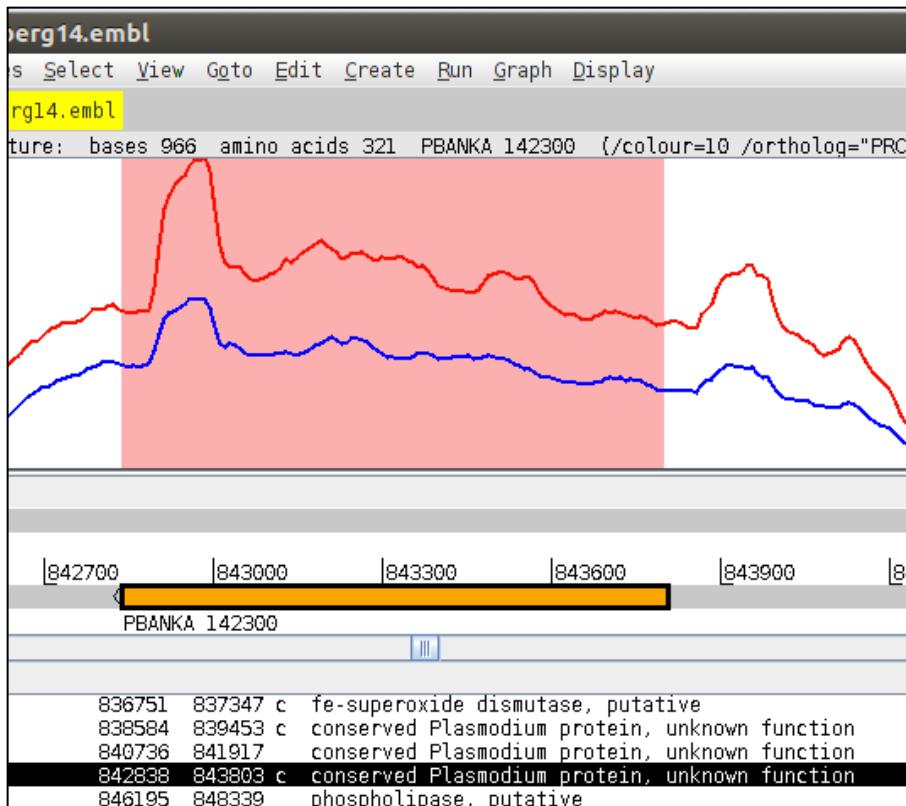


Some definitions

- What does a RNA-Seq read represents?
- Why does that represent expression?
- What does the height of the coverage plot represents?

When is something Differentially expressed?

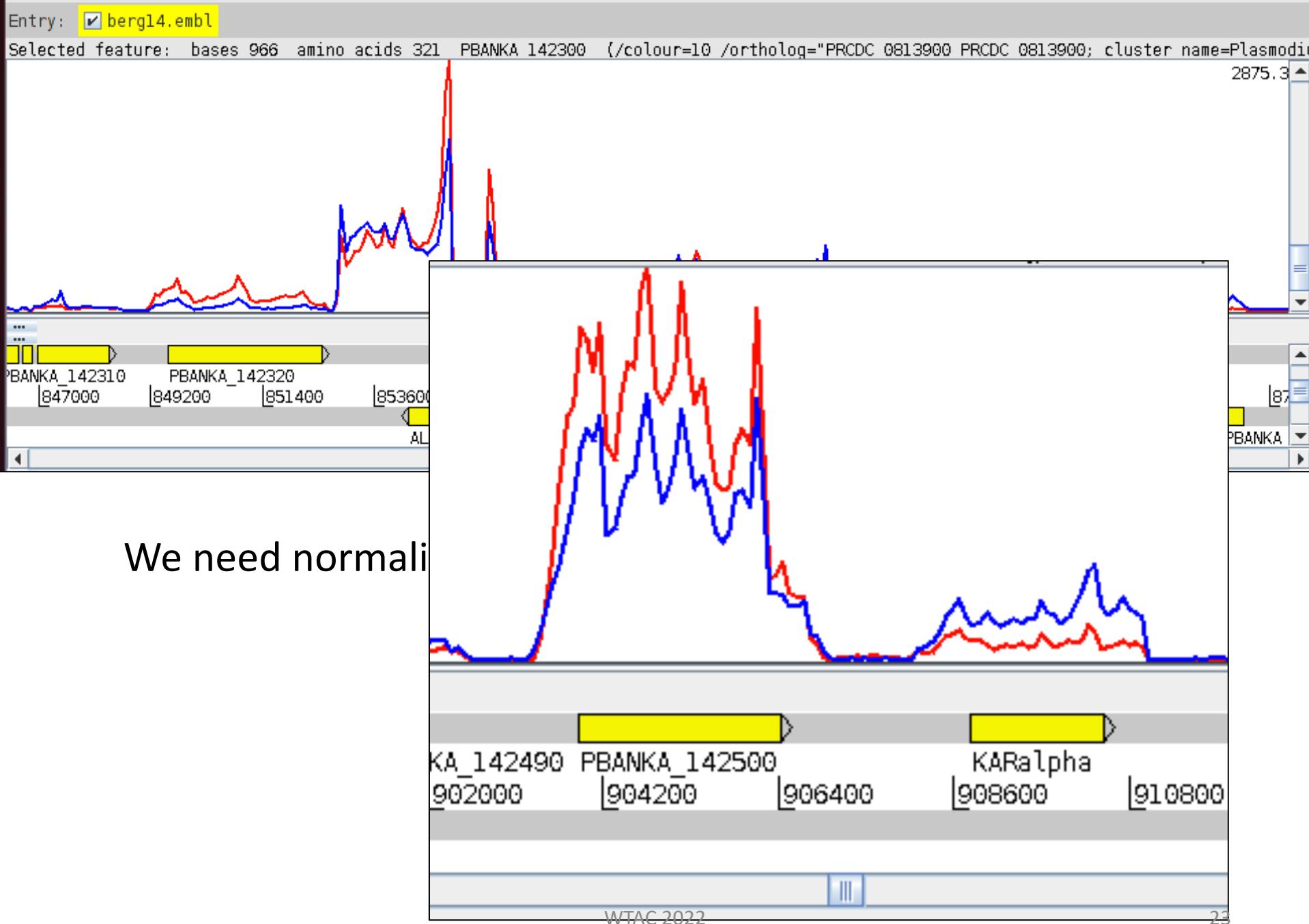




expression pattern of genes in *P. berghei*
 KO - blue
 WT - red

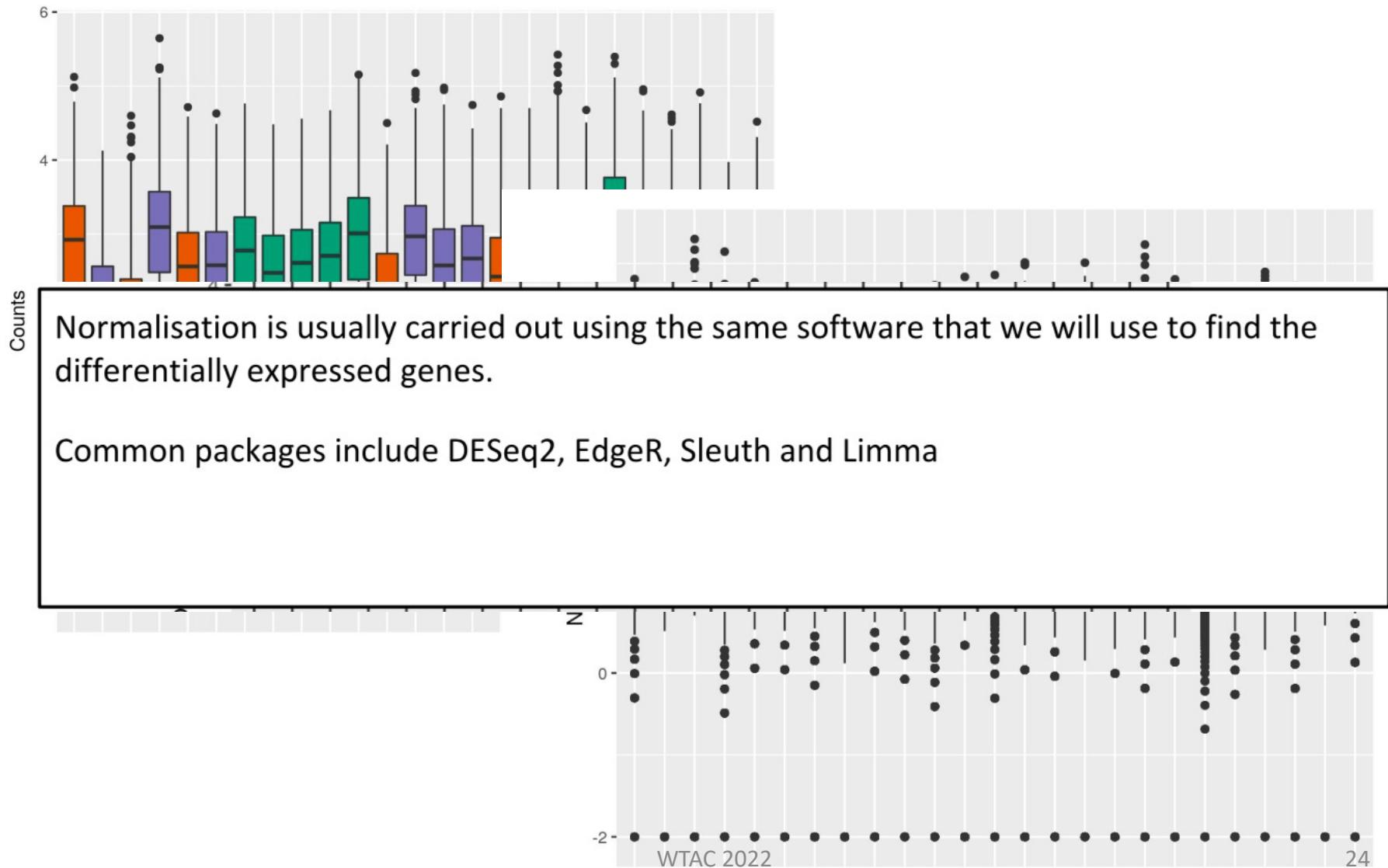
What are you seeing?
 What does the coverage plot represents?

Sample	Total Reads Sequenced
WT	100,000,000
KO	60,000,000

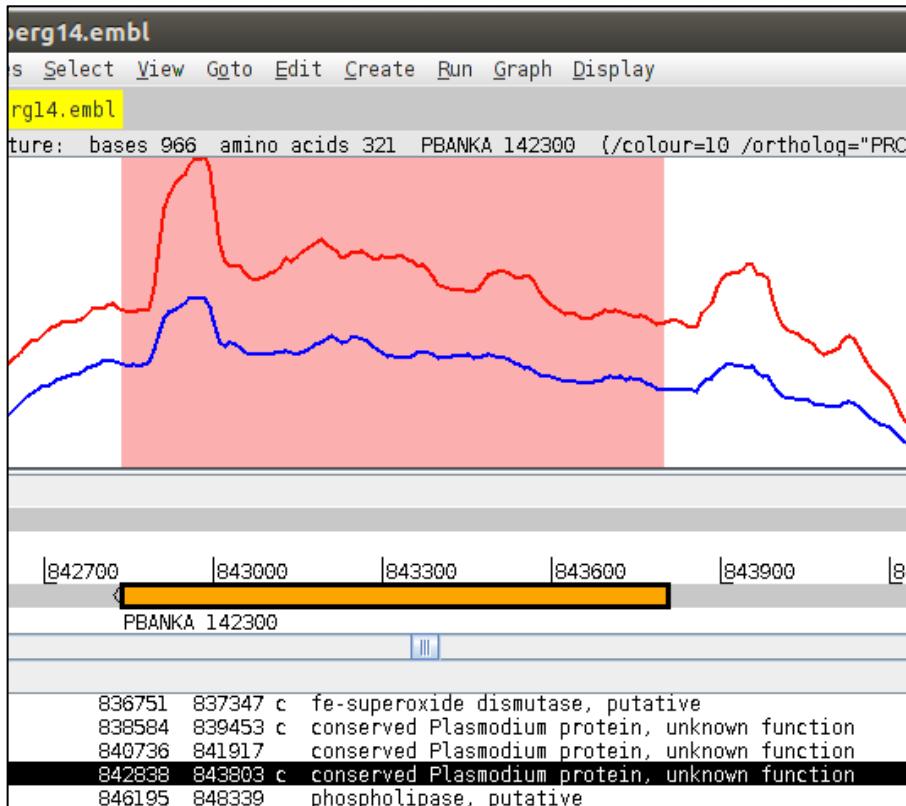


We need normali

Normalisation



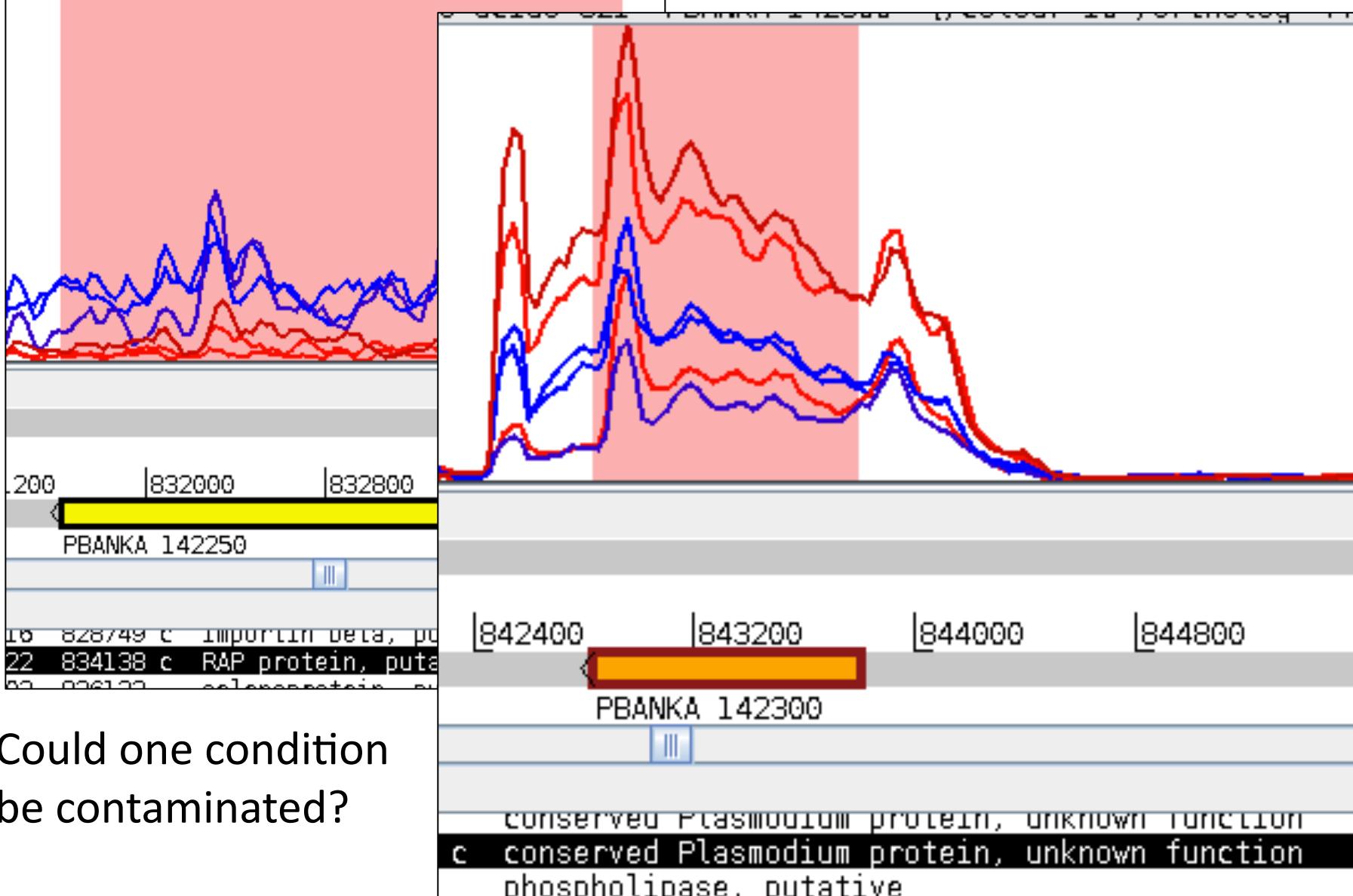
What else do we need?



expression pattern of
genes in *P. berghei*
WT - blue
KO - red

2517 amino acids 838 PBANKA 142250 (/colour)

We need biological replicates!

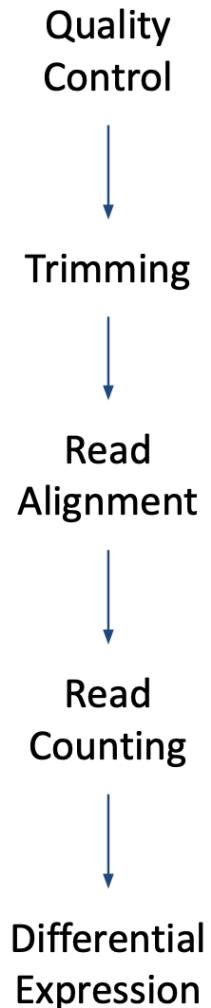


Could one condition
be contaminated?

Differential Expression

After Mapping:

- Count reads mapped to each gene
 - htseq-count
 - featureCount
- Normalise for library size
- Test whether normalised counts differ between conditions
 - DESeq2
 - EdgeR
 - Sleuth



• Comparing Samples

Once data has been normalised, you can compare your conditions

Method is similar to a t-test but uses different tests that are more suited to the data distribution

- Wald test (DESeq2, Sleuth)

- F-test (EdgeR, Limma)

Multiple test correction is essential

Multiple correction

- If you do more test, you need to apply correct for multiple testing!!!
- If you have four conditions A B C D E and do
- A vs B; A vs C; A vs D; A vs E; B vs C; B vs D; B vs E; C vs D; C vs E & D vs E – you have to correct ALL P-values for the 10 amount of test.
- Most conservative method is to multiple so an adjusted P-value of 0.005 becomes 0.05 and is on the boardline of the false discovery rate!
- Tools normally do multiple correction, but not between “experiments” of the user

Differential expression

- Good biological question
- Good sample preparation
- Enough biological replicates (get information through published studies)
- Sequencing -> Quality control, fastqc and visualize the results
- Differential expression (DeSeq2, EdgeR, Sleuth)
- Check FDR, enrichment test (GO)
- Interpretation of the results
- Write the paper

warnings

- RNA-Seq protocols are not always straight forward
- How much RNA can you get?
- Parasite have generally ribosomal RNA you are not interested in
- Are you interesting the expression of genes or RNA's? Poly-A selection
- Expression is more random than genomic, so you will need replicates

Where can I find existing RNA-Seq

Search results for "plasmodium vivax" < ArrayExpress < EMBL-EBI

EMBL-EBI  Services Research Training About us

plasmodium vivax Examples: E-MEXP-31, cancer, p53, Geuvadis Search Advanced

Home Browse Submit Help About ArrayExpress Feedback Login

Filter search results

Search results for plasmodium vivax

+ Show more data from EMBL-EBI

10 experiments

Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views	Atlas
E-GEOD-61252	New insights into the Plasmodium vivax transcriptome using RNA-seq	RNA-seq of coding RNA	Plasmodium vivax	19	11/02/2016			49	-
E-GEOD-67469	Transcription Profiling of Malaria-Naïve and Semi-Immune Colombian Volunteers in a Plasmodium vivax Sporozoite Challenge [RT-PCR]	transcription profiling by RT-PCR	Homo sapiens	87	14/07/2015		-	25	-
E-GEOD-67184	Transcription Profiling of Malaria-Naïve and Semi-Immune Colombian Volunteers in a Plasmodium vivax Sporozoite Challenge [RNA-Seq]	RNA-seq of coding RNA	Homo sapiens	24	14/07/2015	-		68	-
E-GEOD-45165	Natural antisense transcripts in Plasmodium vivax clinical isolates	transcription profiling by array	Plasmodium vivax	2	16/01/2014		-	200	-
E-GEOD-33333	Characterization and gene expression analysis of the cir multi-gene family of Plasmodium chabaudi chabaudi (AS)	transcription profiling by array	Plasmodium chabaudi chabaudi	12	31/05/2012			89	-
E-GEOD-12174	Identification of Plasmodium vivax genes whose expression is spleen-dependent	transcription profiling by array	Plasmodium vivax	5	10/07/2011			142	-
E-MTAB-636	RNA-Seq of Plasmodium yoelii wild type and mutant with deleted one of the RH family genes	RNA-seq of coding RNA	Plasmodium yoelii	5	25/04/2011	-		187	-
E-GEOD-23982	Whole-genome sequencing and microarray analysis of ex vivo Plasmodium vivax reveal selective pressure on putative drug resistance genes	comparative genomic hybridization by array	Plasmodium vivax	3	02/10/2010		-	53	-

VEuPathDB

Fold Change Percentile

Identify Genes based on *T. cruzi* CL Brener Esmeraldo-like Transcriptomes of Four Life-Cycle Stages Microarray (fold change)

[Tutorial](#) [Help](#)

For the Experiment

Transcriptomes of Four Life-Cycle Stages tcruCLBrenerEsmeraldo-like

return protein coding Genes
that are up-regulated
with a Fold change >= 2.0

between each gene's expression value
in the following Reference Samples

amastigotes
 trypomastigotes
 epimastigotes
 metacyclics
[select all](#) [clear all](#)

and its expression value
in the following Comparison Samples

amastigotes
 trypomastigotes
 epimastigotes
 metacyclics
[select all](#) [clear all](#)

Example showing one gene that would meet search criteria
(Dots represent this gene's expression values for selected samples)

Up-regulated

Expression

Comparison

Reference

Reference Comparison
Samples Samples

You are searching for genes that are up-regulated between one reference sample and one comparison sample.

For each gene, the search calculates:

$$\text{fold change} = \frac{\text{comparison expression value}}{\text{reference expression value}}$$

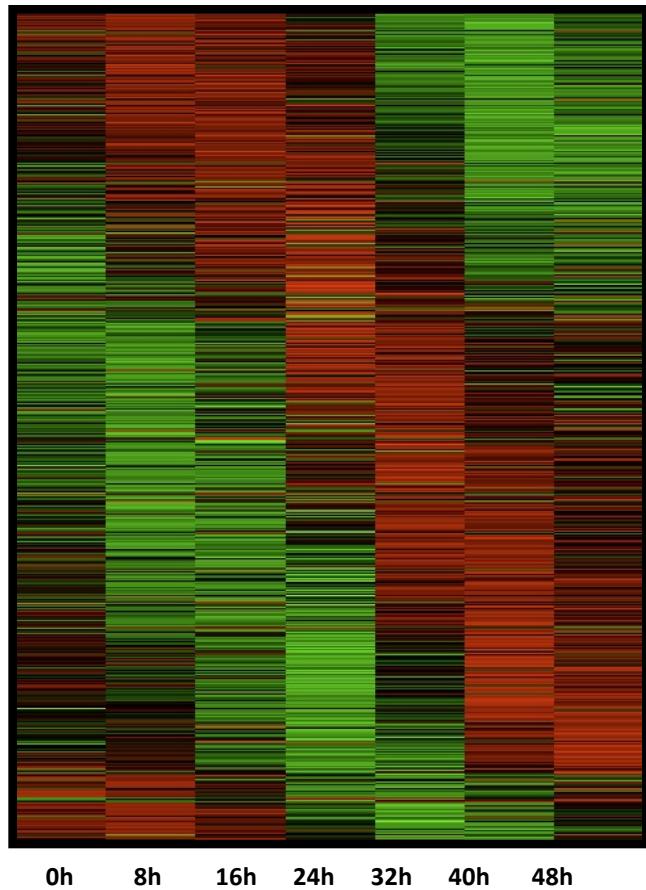
and returns genes when fold change >= 2.0.

See the detailed help for this search.

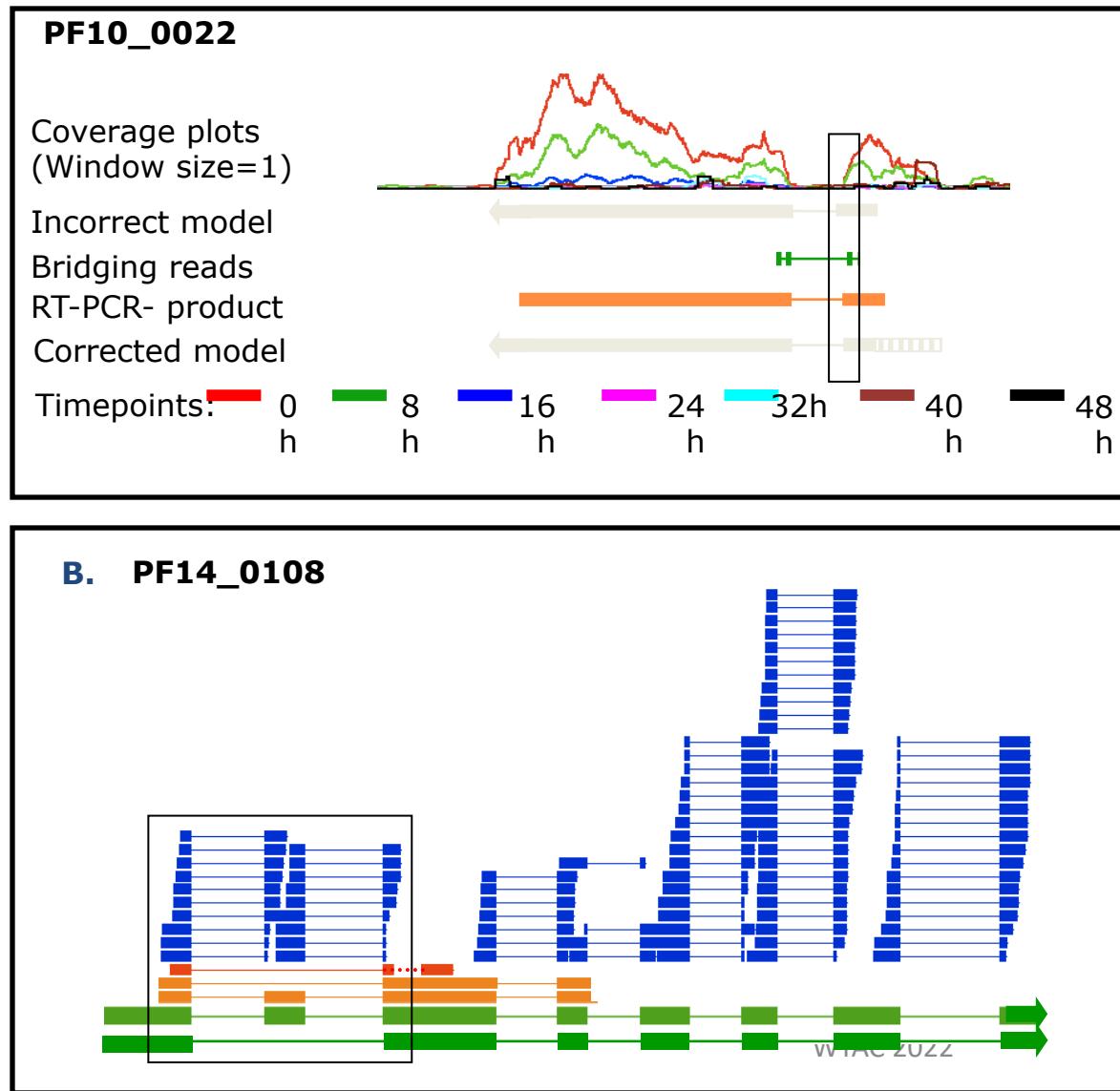
Examples

RNA-Seq in *P. falciparum*

A. IDC RNA-Seq data



Otto et al.
Mol. Microbiology 2010



Knock out of gene

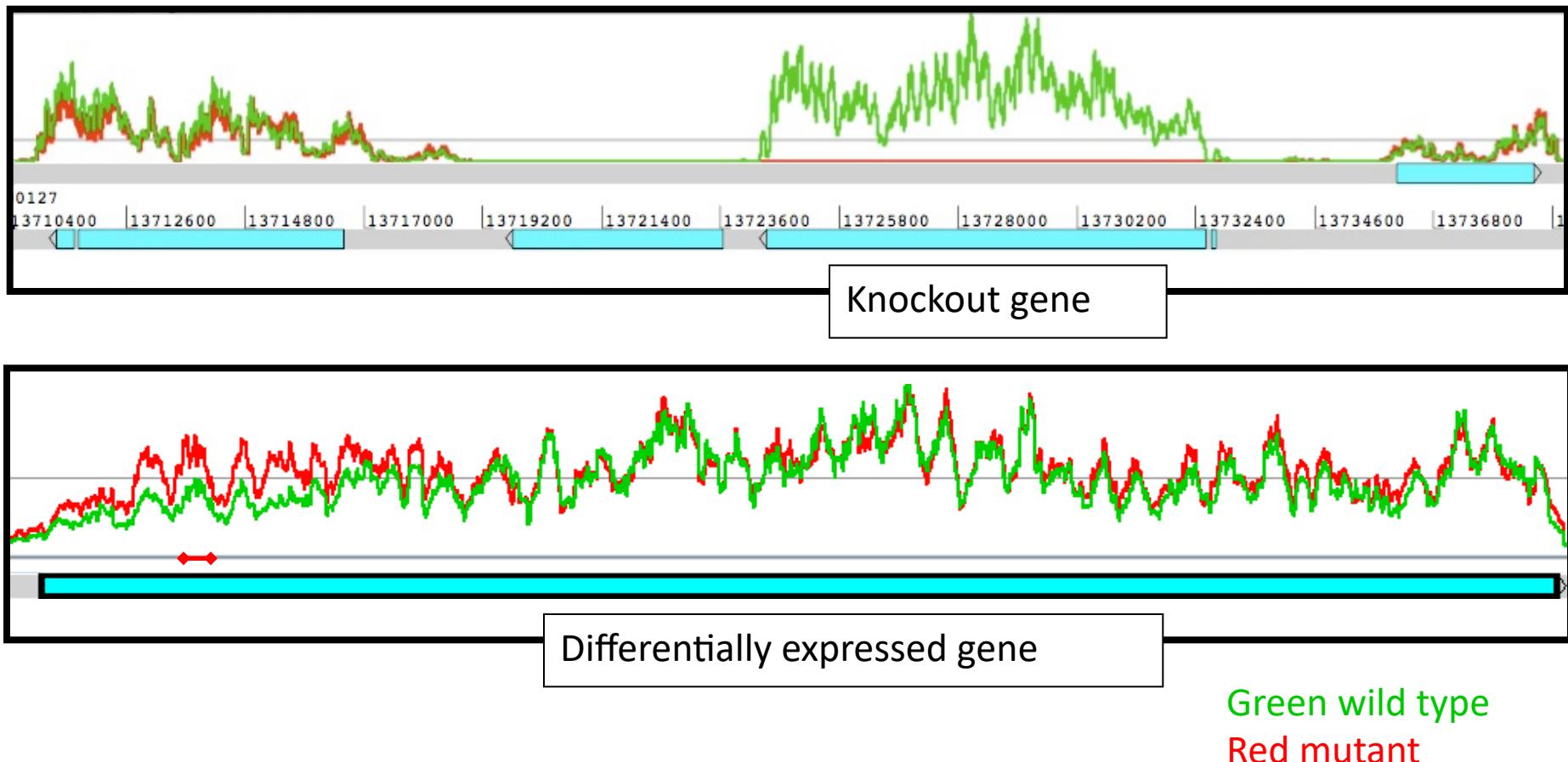
- Invasion protein
- Gene not essential
- How will the expression profile change

Targeted Disruption of *py235ebp-1*: Invasion of Erythrocytes by *Plasmodium yoelii* Using an Alternative Py235 Erythrocyte Binding Protein

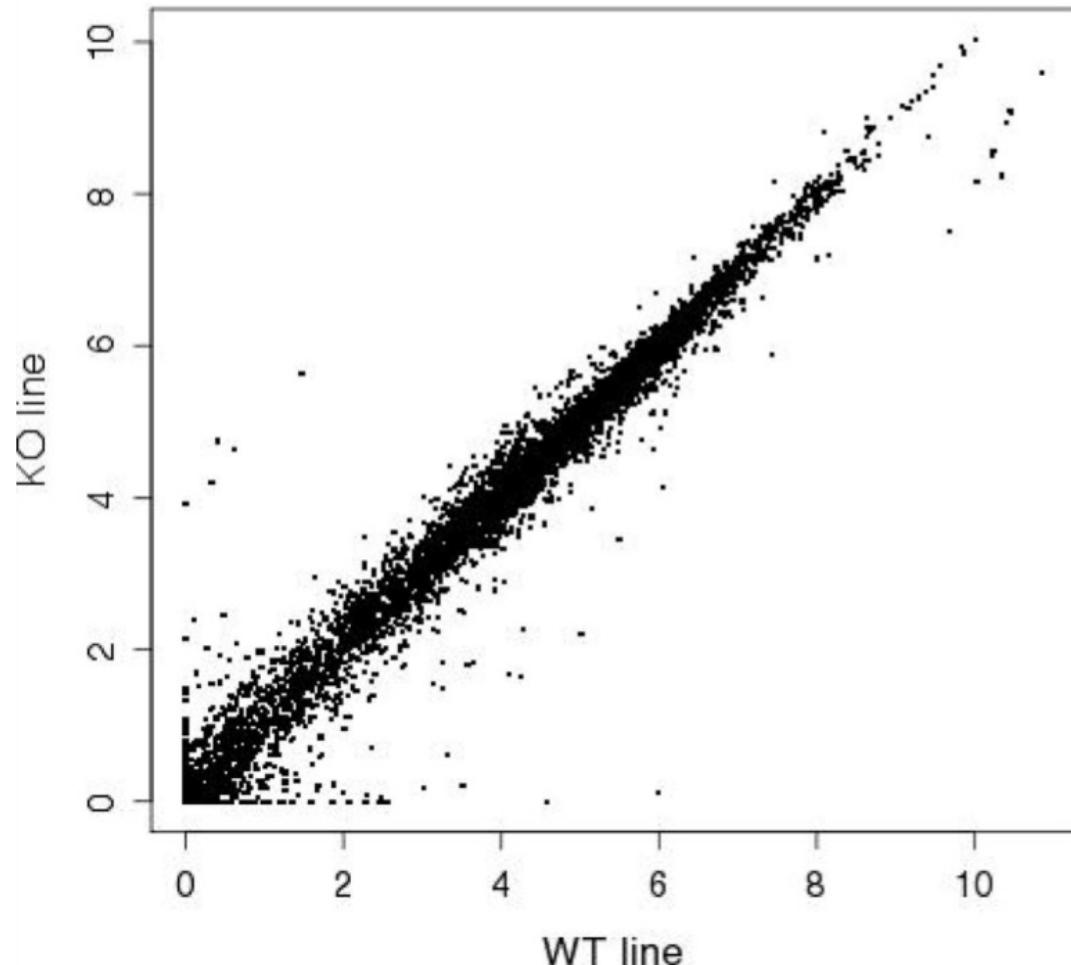
Solabomi A. Ogun^{1*}, Rita Tewari³, Thomas D. Otto⁴, Steven A. Howell², Ellen Knuepfer¹, Deirdre A. Cunningham¹, Zhengyao Xu³, Arnab Pain^{4,5}, Anthony A. Holder^{1*}

PLoS Pathogens

Differential expression after knockout



Correlation



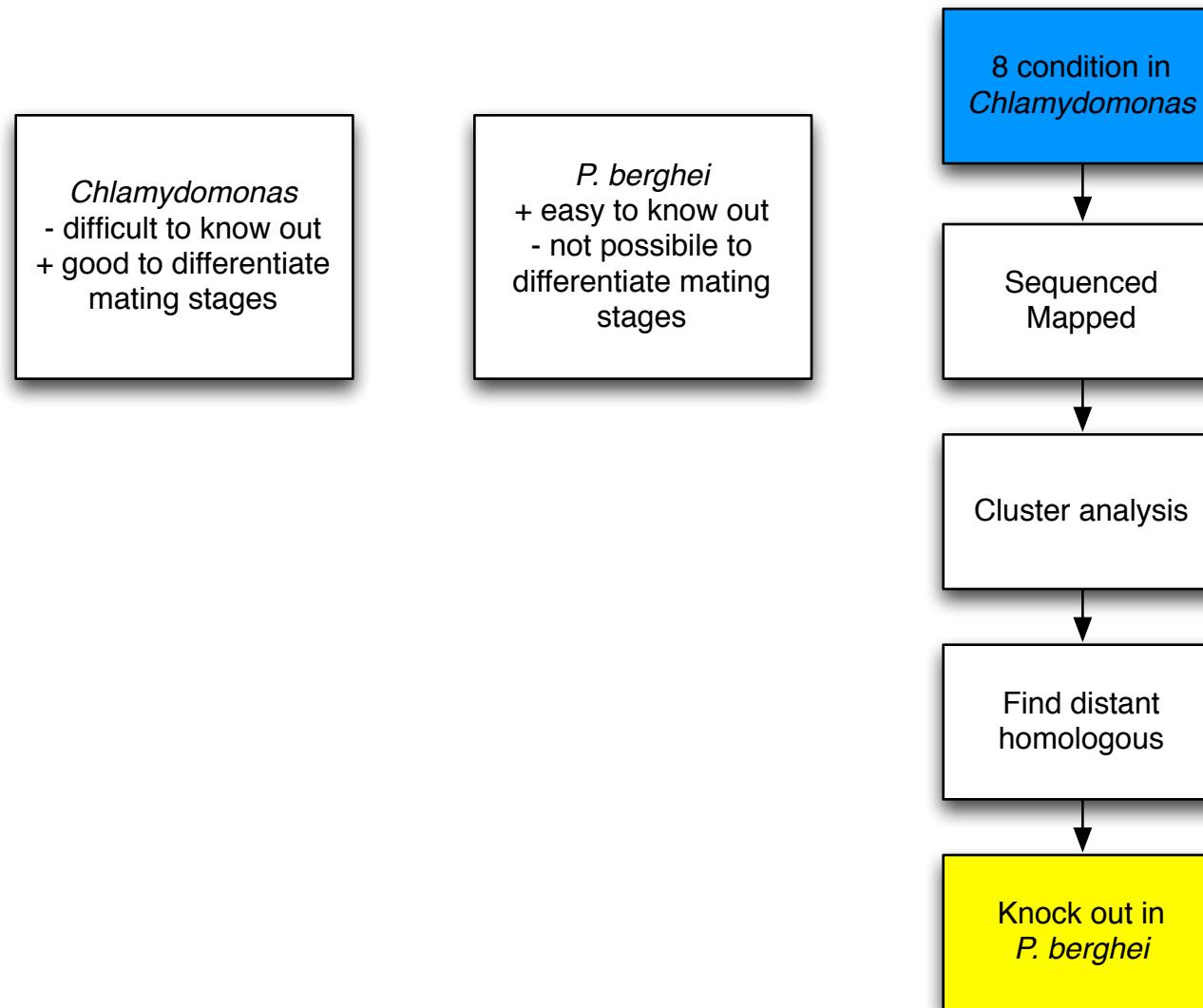
What does it tell us?

Is it differential expressed?

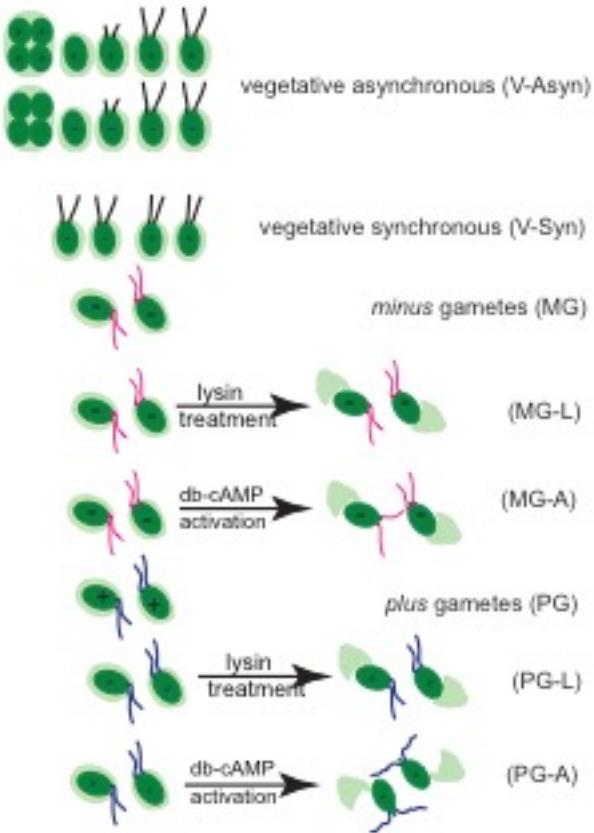
Is it statistical meaningful?

Better tools, if you have biological replicates, like DESEQ or EDGER

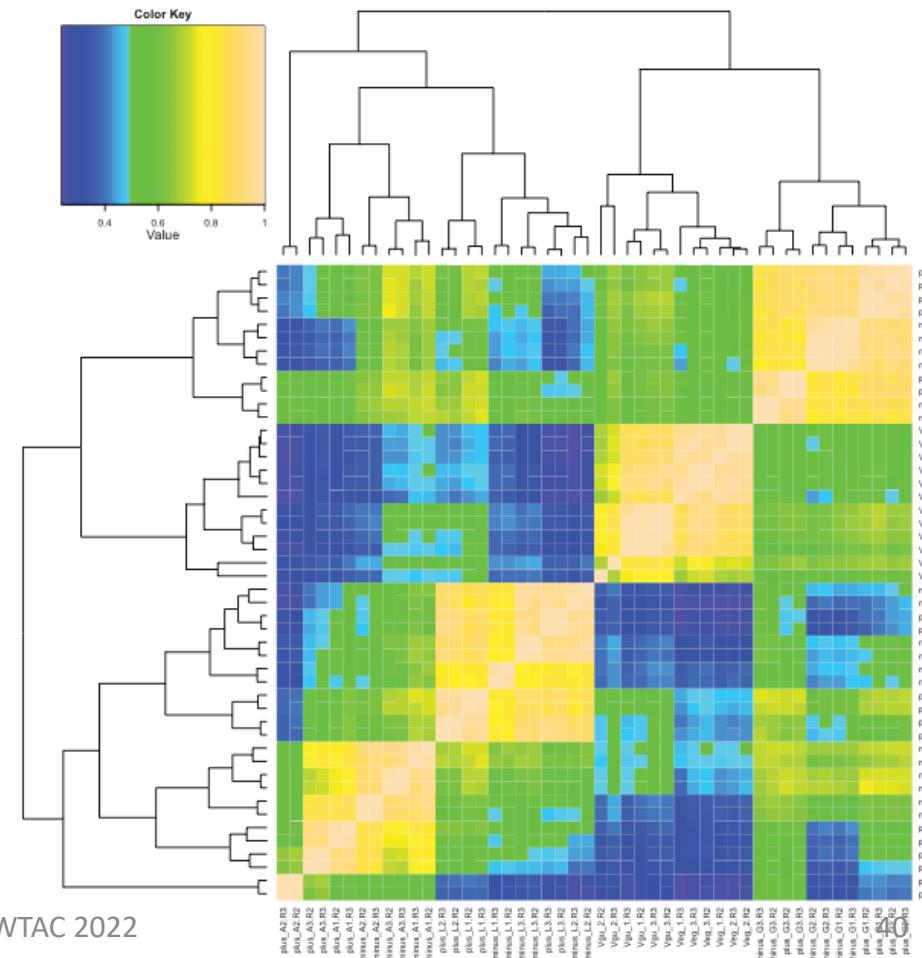
Comparative genomics of sexual reproduction in *Chlamydomonas* and *Plasmodium*



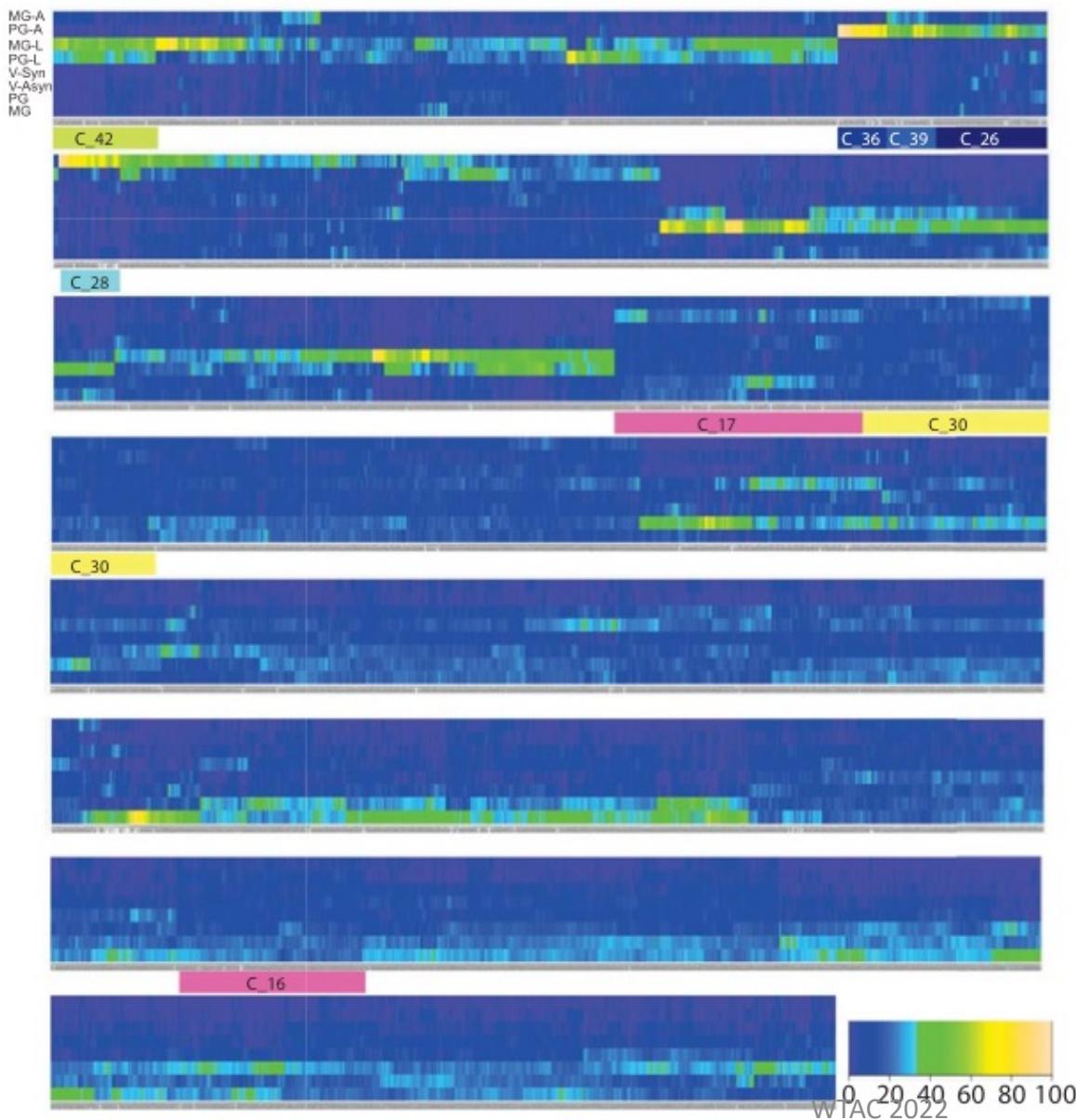
Conditions



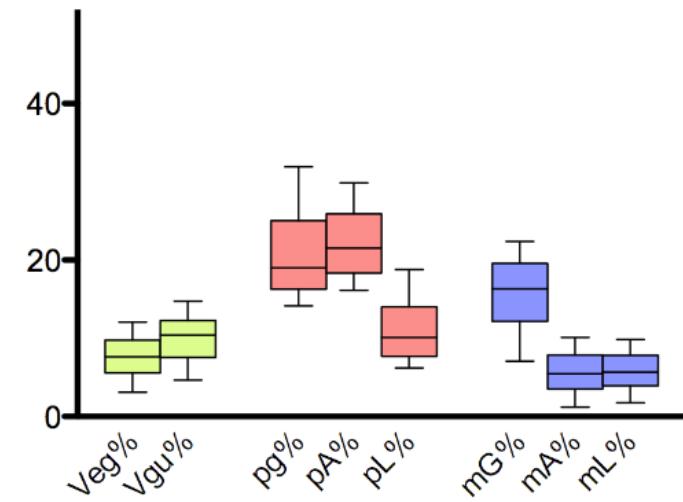
Correlation



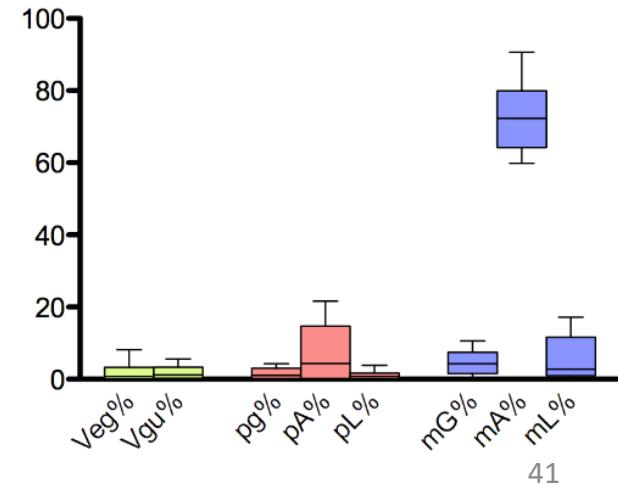
Cluster analysis



Cluster 17: 344 genes

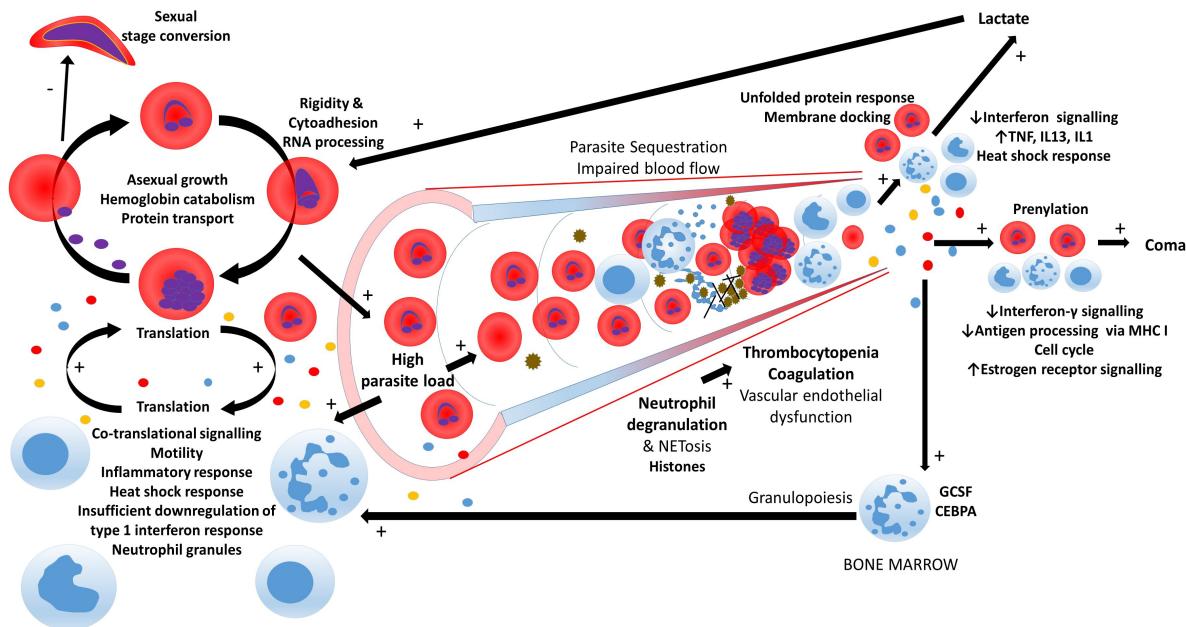


Cluster 28: 85 genes



Host-Parasite Dual RNA-Seq

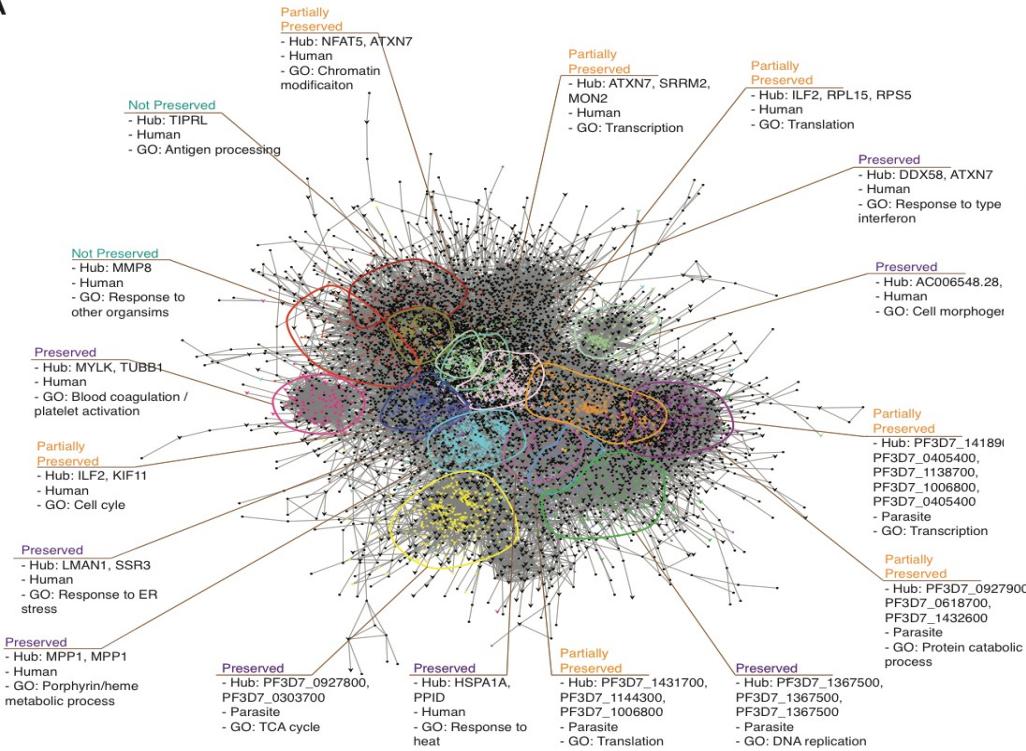
- Led by Aubrey Cunnington, Imperial College, London
- 46 Sample, severe and non severe from kids



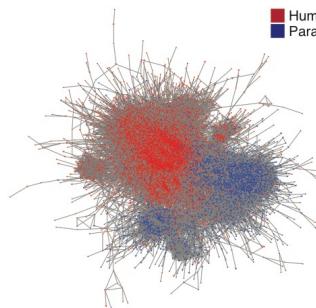
Lee J, science Translational medicine, 2018

Co-expression Network

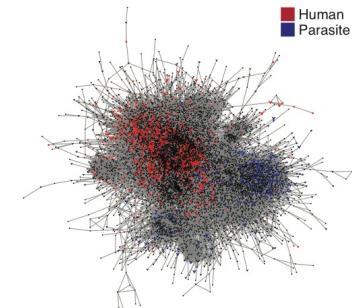
A



B



C



<https://pubmed.ncbi.nlm.nih.gov/29950443/>

Outlook

- Now single cell RNA-Seq exists
- Spatial single cell transcriptomics is coming

scRNA-Seq advances

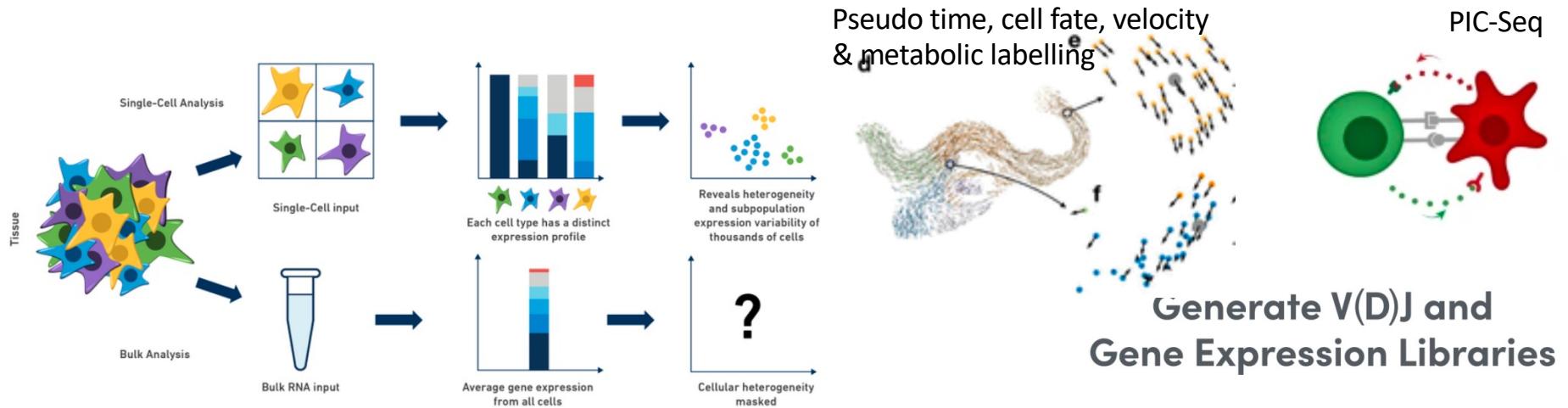
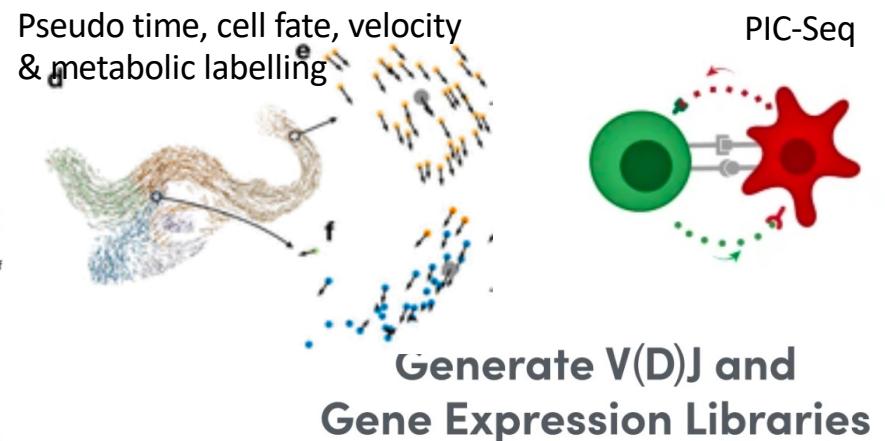
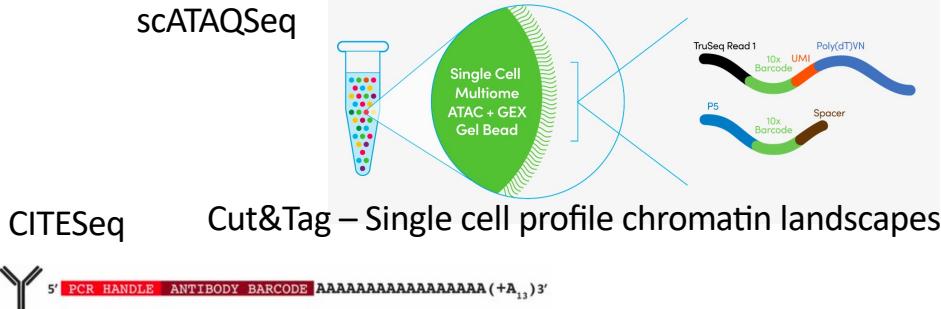
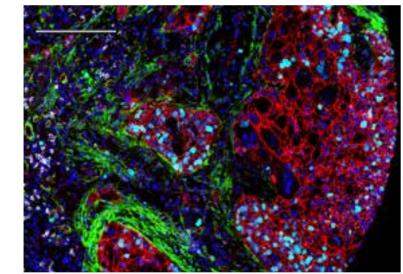
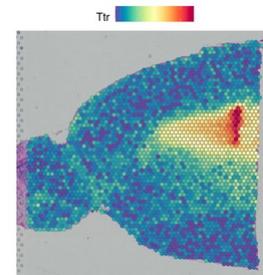


Figure 1. Single-cell RNA-seq reveals cellular heterogeneity that is masked by bulk RNA-seq methods.



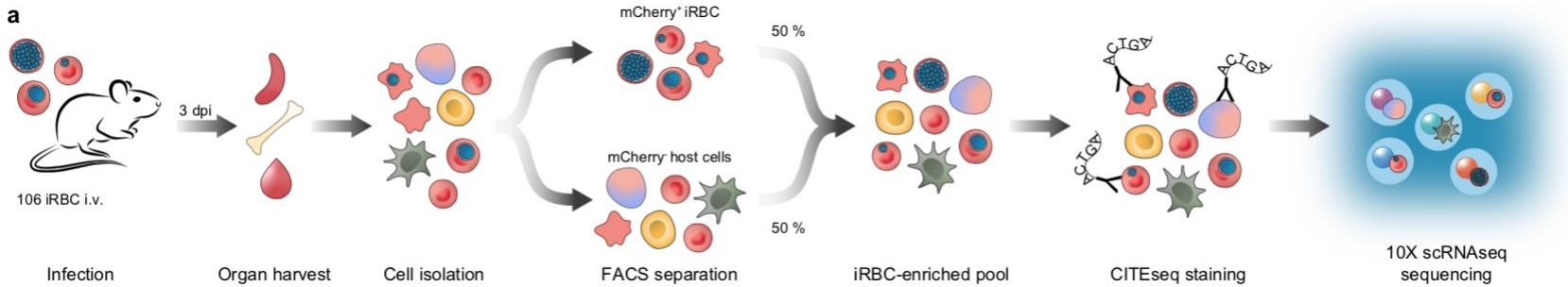
Perturb-Seq – single cell CRISPR/Cas9

10x Visium (and others)
Spatial transcriptomics



imaging mass cytometry
(Hyperion)

Host parasite interaction in Malaria



Franziska Hentzschel

Science Advances

Current Issue First release papers Archive About [Submit](#)

HOME > SCIENCE ADVANCES > VOL. 8, NO. 17 > HOST CELL MATURATION MODULATES PARASITE INVASION AND SEXUAL DIFFERENTIATION IN PLASMODIUM BERGHEI

[RESEARCH ARTICLE](#) | MICROBIOLOGY

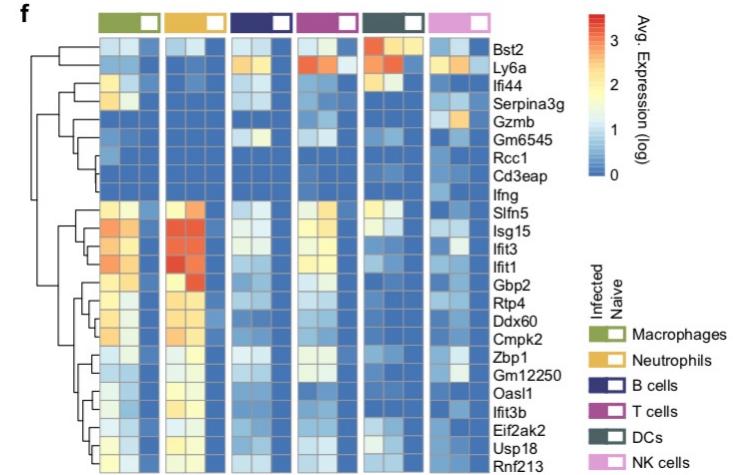
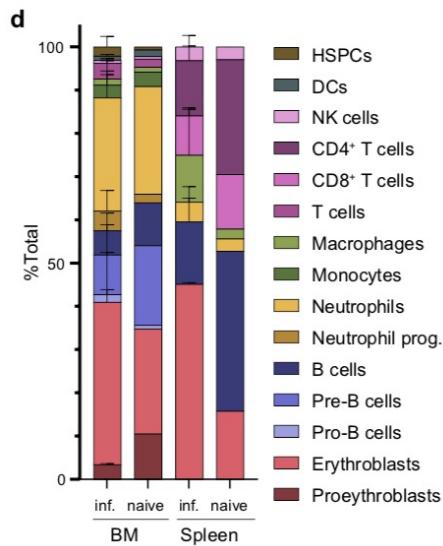
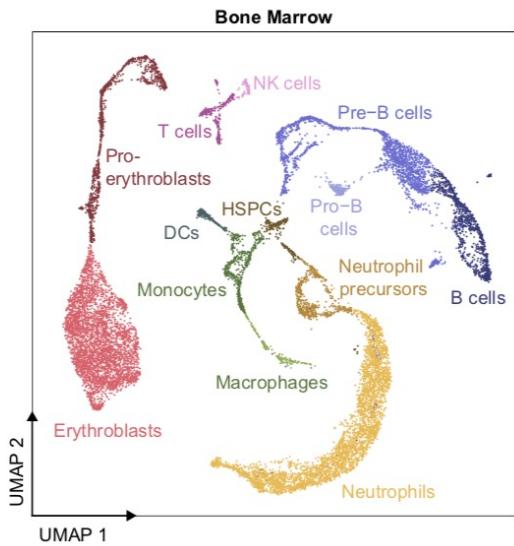


Host cell maturation modulates parasite invasion and sexual differentiation in *Plasmodium berghei*

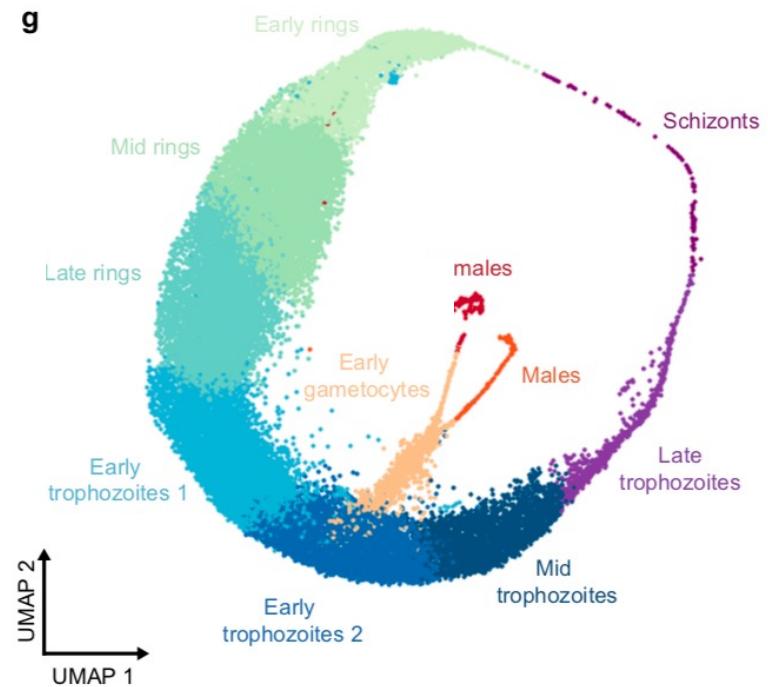
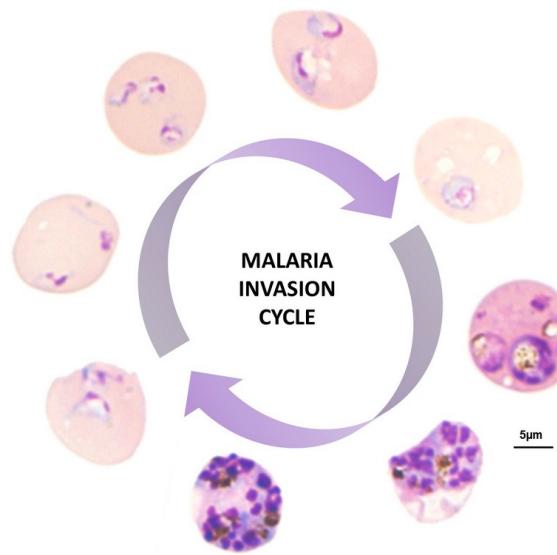
FRANZISKA HENTZSCHEL , MATTHEW P. GIBBINS , CHARALAMPOS ATTIPA , DARIO BERALDI , CHRISTOPHER A. MOXON , THOMAS D. OTTO , AND MATTHIAS MARTI [Authors Info &](#)

Affiliations

Host parasite interaction in Malaria

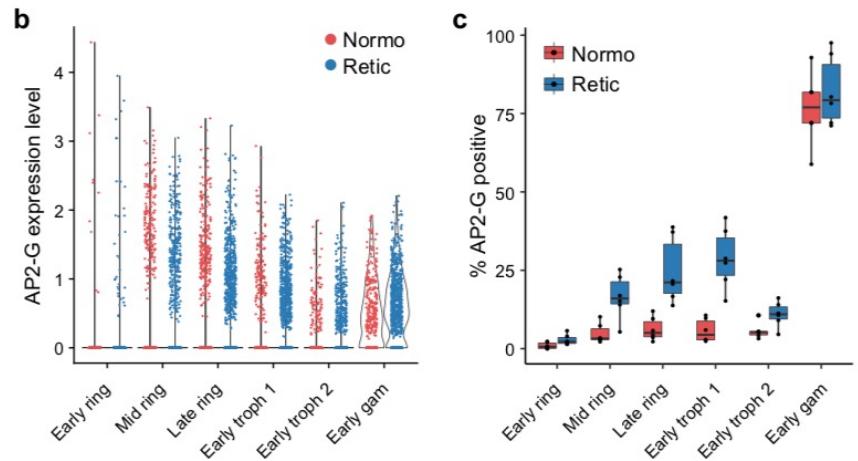
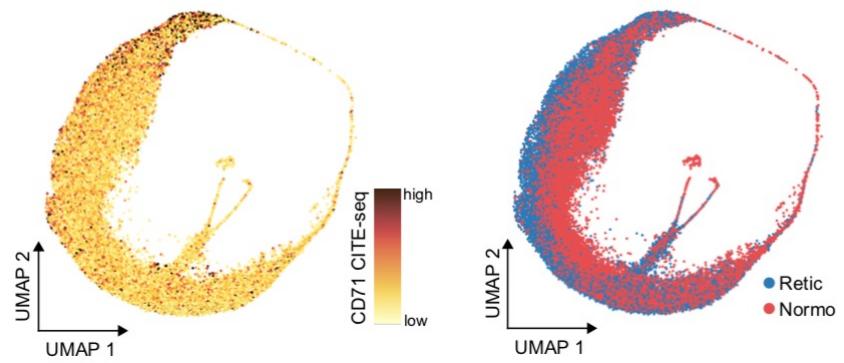
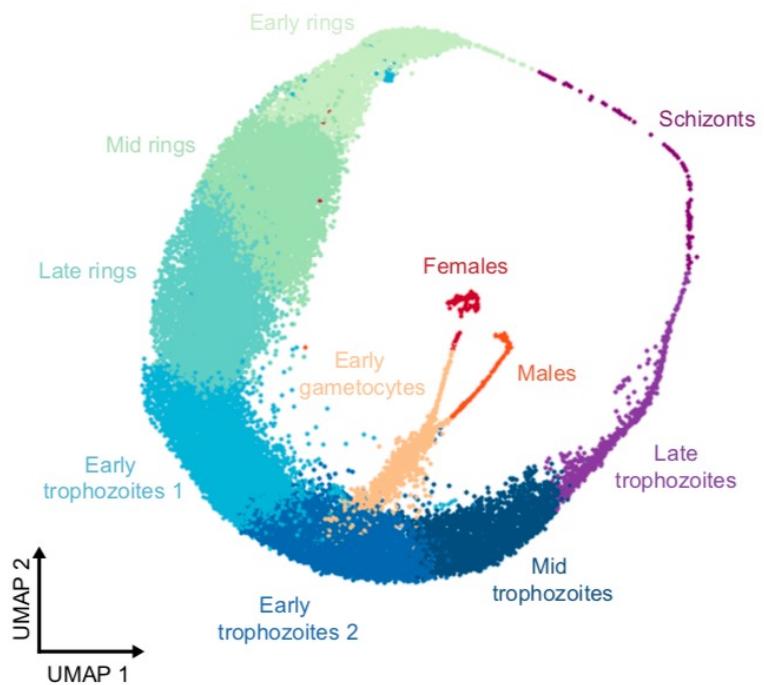


Invasion phenotype



P. berghei sexual development occurs preferentially in reticulocytes

Invasion phenotype



P. berghei sexual development occurs preferentially in reticulocytes

Henschel *et al*, Science advances 2022

Conclusions

- RNA-Seq is a powerful tool
- Improve Annotation / New functional elements
- Differential expression / Cluster analysis to find genes changing expression due to different condition
- different new methods to combine the data (integrated biology)