ZENCODE-ITN

Advanced RNA-seq training

Time course experiment analysis

Damir Baranasic

1st December 2017

RNA-seq

whole transcriptome assessment

novel and known transcripts

allele specific expression

gene fusion

lncRNA

eRNA

alternatively spliced variants

transcripts quantification

differential expression

Time course experiments

development and growth

reaction to a treatment/condition over time

single time series

one condition

all time points compared to the first one (control)

multi time series

several conditions simultaneously

controls are sampled over time with the samples

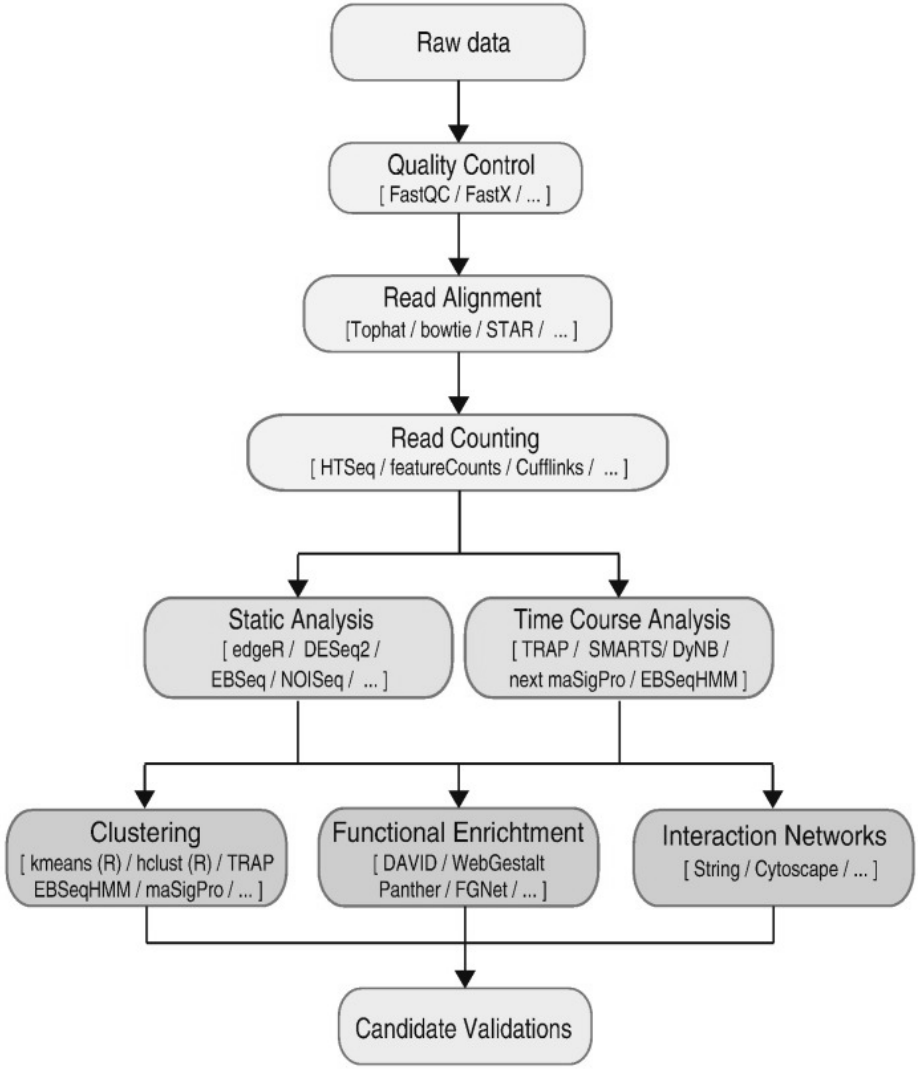
periodic and cyclic time series

single or multiple conditions

reoccurring expression patterns and their difference between conditions

complex >> a lot of samples and synchronization needed

Analysis workflow



similar to
static RNA-seq experiments

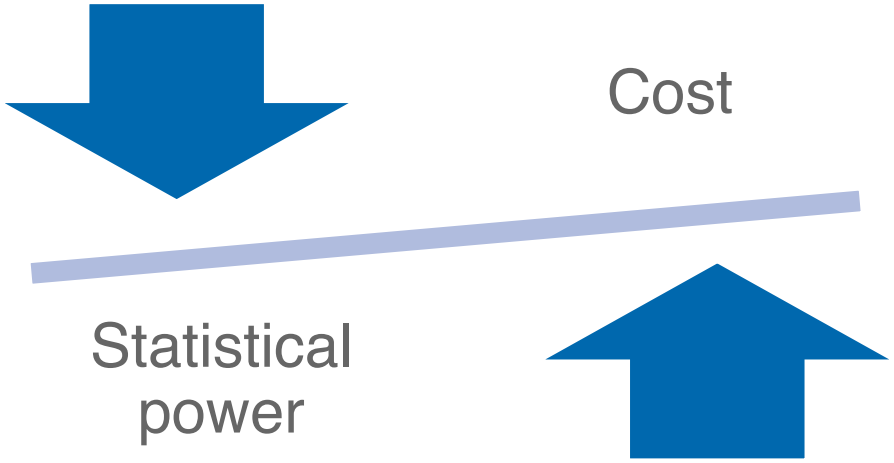
Experimental design

critical

number of time points

number of replicates

there are tools to estimate this parameters, but they don't consider multi-factor experiments



when in doubt

more replicates better than greater sequencing depth

bad design

👎 statistical power 👆 number of false positives

Data analysis

static tools

sequencing depth and library size

batch effect >> protocol

sequencing platform

technical variability

time course

do not consider correlation between neighbouring time points

Choosing the right method

questions to address

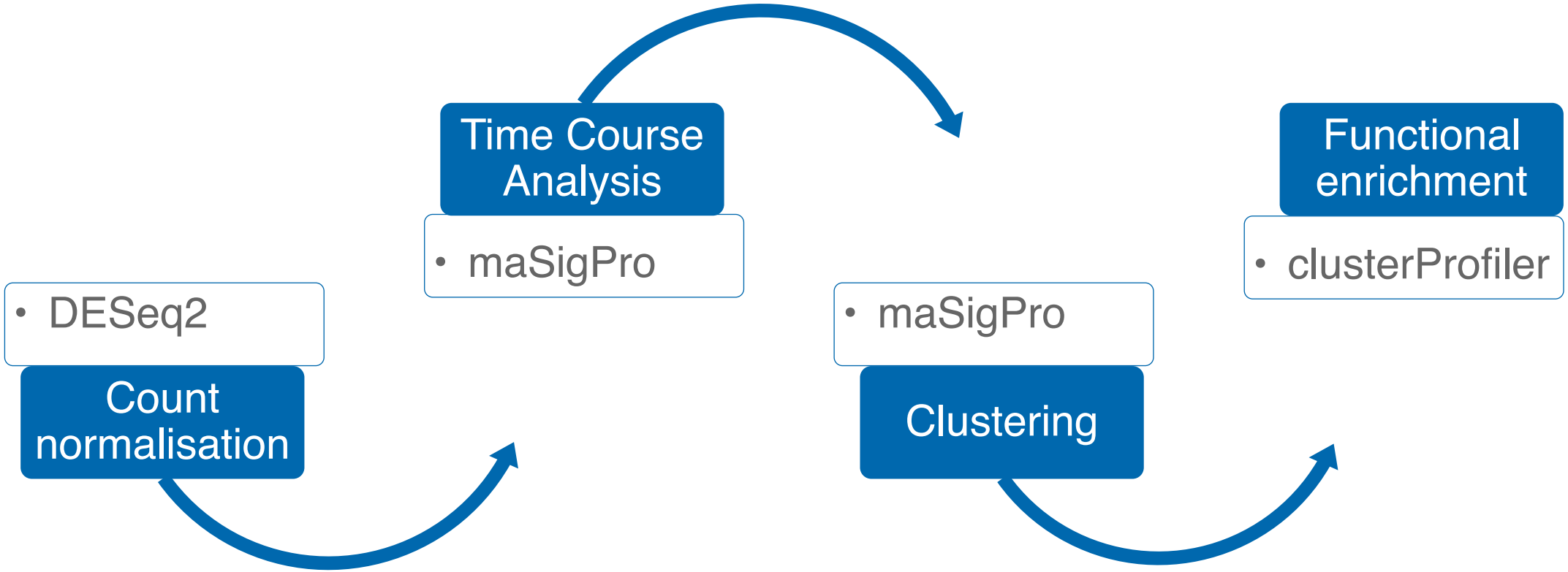
number of replicates

experimental design

two-way or multi factor

differential expression of RNA isoforms

Training steps



Workflow

data normalisation with DESeq2

size factor correction drawn from the negative binomial distribution

maSigPro

GLM with negative binomial distribution as a linking function

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \beta_3 z_{1i} + \beta_4 t_i z_{1i} + \beta_5 t_i^2 z_{1i} + \varepsilon_i$$

functional enrichment with clusterProfiler

hypergeometric distribution testing

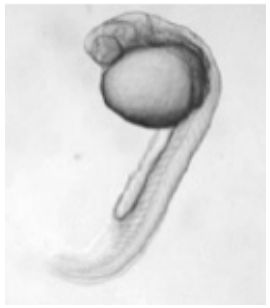
finds functional terms occurring significantly more thann expected

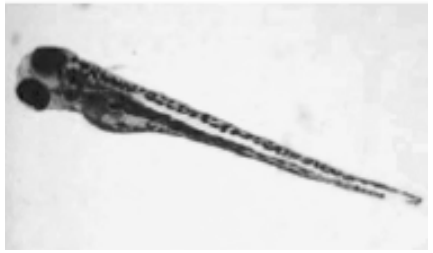
ZENCODE-ITN

Data used in this training

Christelle Etard, Olivier Armant, Urmas Roostalu, Victor Gourain, Marco Ferg and Uwe Strähle.
Loss of function of myosin chaperones triggers Hsf1-mediated transcriptional response in skeletal muscle cells.
Genome Biology. 2015; 16:267 https://doi.org/10.1186/s13059-015-0825-8

`RNA-seq_Strahle_Lab_0005AS.<SequencingID>.USERvgourain.R.ReadsPerGene.out.tab`





Hpf	24	48	72
wt	DCD001548SQ	DCD001546SQ	DCD001547SQ
	DCD001559SQ	DCD001558SQ	DCD001545SQ
unc45b	DCD001560SQ	DCD001564SQ	DCD001565SQ
	DCD001554SQ	DCD001555SQ	DCD001551SQ

All the libraries were:

unstranded
paired-ended
sequenced on Illumina HiSeq 2000
producing 50 bp long reads.

References

[1] Daniel Spies and Constance Ciaudo.
Dynamics in Transcriptomics: Advancements in RNA-seq Time Course and Downstream Analysis. *Comput Struct Biotechnol J.* 2015; 13: 469–477.

[2] Michael I Love, Wolfgang Huber and Simon Anders.
Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology.* 2014; 15: 550.

[3] María José Nueda, Sonia Tarazona and Ana Conesa.
Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics*. 2014; 30(18): 2598–2602.

[4] Guangchuang Yu, Li-Gen Wang, Yanyan Han and Qing-Yu He.
clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*. 2012; 16(5): 284-287.