bulkAnalyseR: An accessible, interactive pipeline for analysing and sharing bulk multi-modal sequencing data

Eleanor Williams

Wellcome-MRC Cambridge Stem Cell Institute

bulkAnalyseR aims to facilitate downstream analysis without sacrificing customisation

- Analysing RNA-seq data requires programming expertise to manipulate the data into the correct format for different tools

- If you don't have that knowledge, you need bioinformatics support or are reliant on commercial tools which can be expensive and inflexible

- Even if you have bioinformatics support, feedback loops can take a long time and it's difficult to dig into the results you are seeing

- Therefore, we need an open-source, user-friendly pipeline which provides interactivity, flexibility and customisation

	bulkAnalyseR	Searchlight	BioJupies	VIPER	iDEP	DEBrowser	Degust	GENAVi	DEApp	DEIVA
Pre-processing	✓	✗	✗	✓	✓	✓	✗	✓	✓	✗
Quality Checks	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗
Differential expression	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗
Enrichment analysis	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓
Pattern identification	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
Functional analysis (GRN)	✓	✓*	✗	✗	✗	✗	✗	✗	✗	✗
Multi-modal integration	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
Hosted/shared online	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓
Beginner friendly	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓
Parameter adjustment	✓	✗	✗	✗	✓	✓	✓	✗	✓	✓

*Single static representation without comparison across conditions

Our modular app-based approach provides an easy, out-of-the box solution

In only 2 lines of code, you can create an app which does everything in the workflow

```
#run preprocessing
expression.matrix.preproc <- preprocessExpressionMatrix(expression.matrix)

#create shiny app
shiny.dir <- "shiny_Yang2019"
generateShinyApp(
  expression.matrix = expression.matrix.preproc,
  metadata = metadata,
  shiny.dir = shiny.dir,
  app.title = "Shiny app for three timepoints from the Yang 2019 data",
  organism = "mmusculus",
  org.db = "org.Mm.eg.db"
)

#run shiny app
shiny::runApp(shiny.dir)
```
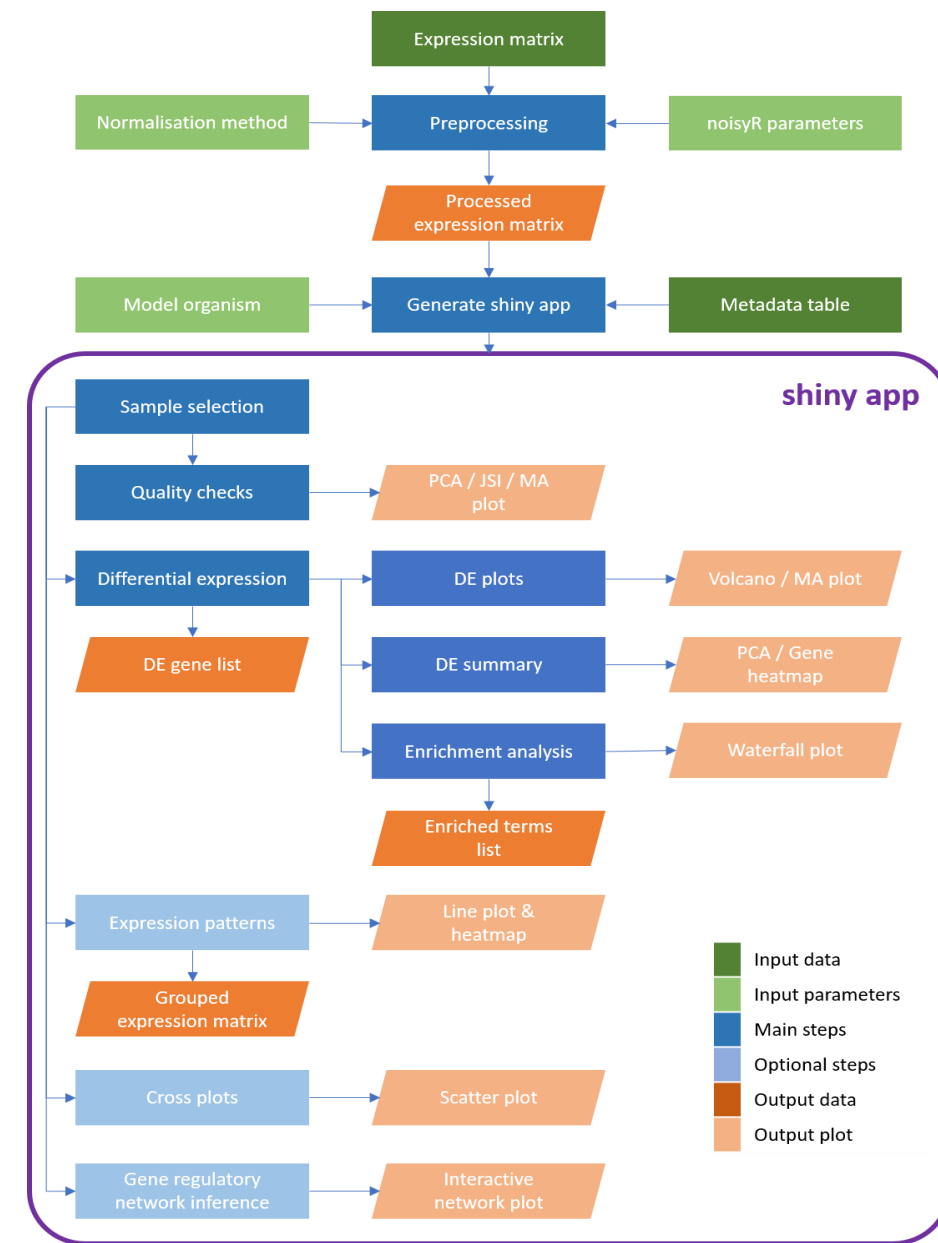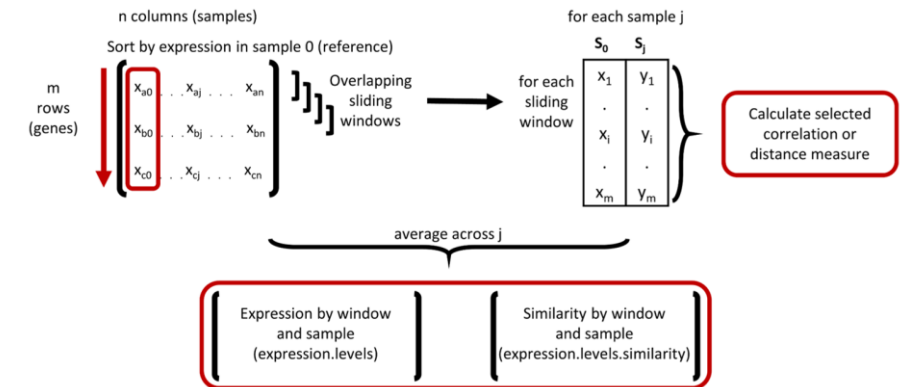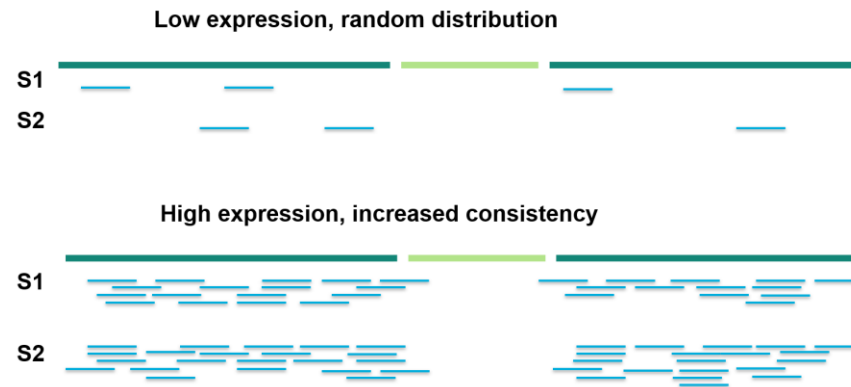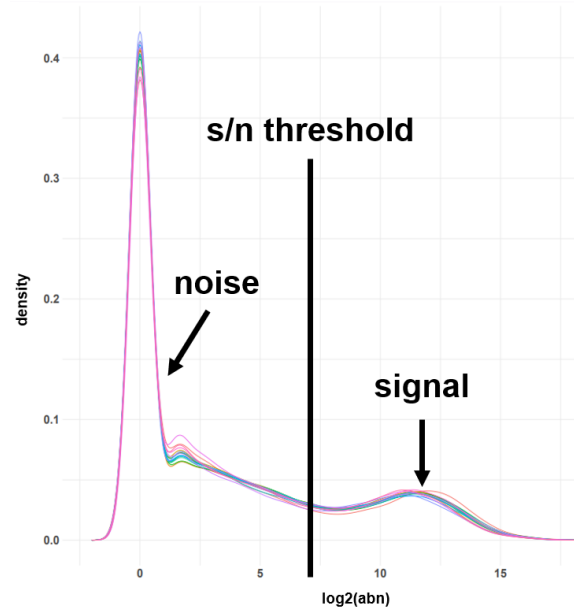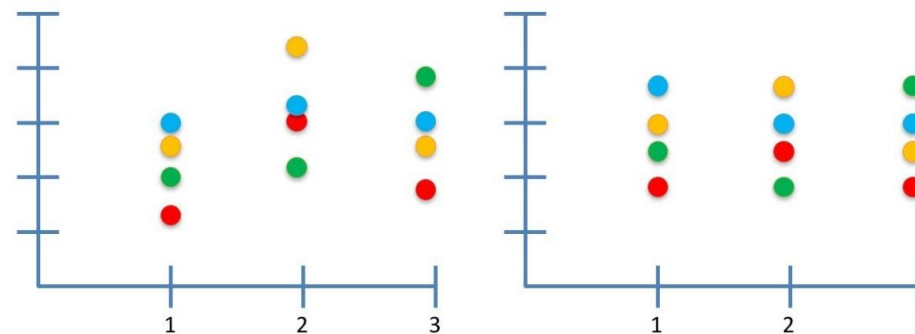
Noise removal and normalization are essential

The noisyR pipeline is used to minimise technical noise

A variety of normalization options is provided, including parametric and non-parametric methods

Using the shiny module framework increases simplicity and versatility

- App is standalone and can be hosted / shared once data is published so the community can interact with the data themselves

- Modules are independent so a full set of tabs can be created for multiple modalities

- Modifications & additions are straight-forward

- Outputs are adjusted interactively

UNIVERSITY OF CAMBRIDGE
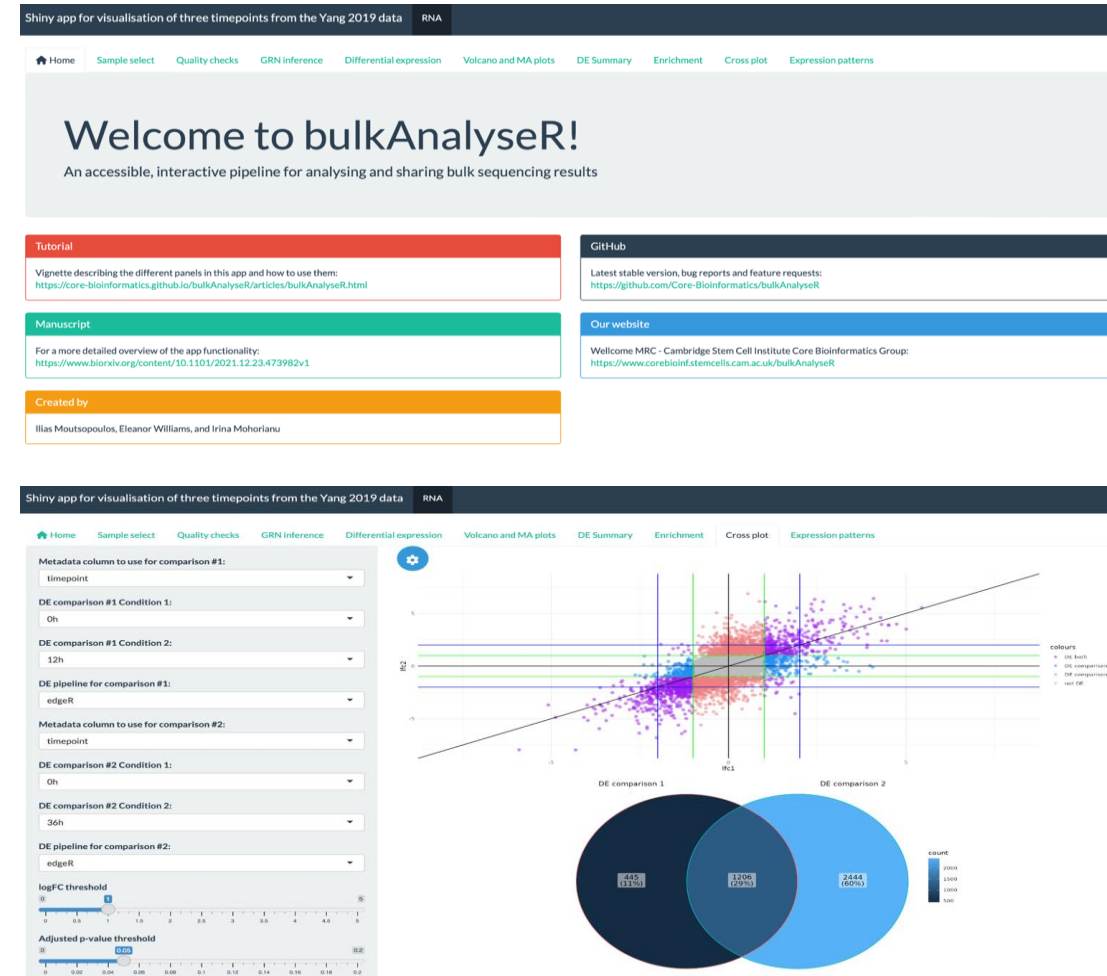
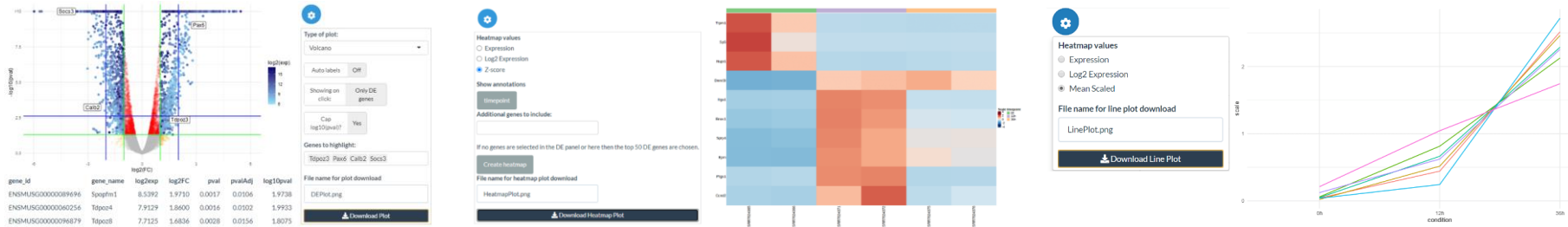# The app functionality includes several standard bulk RNA-seq data analysis steps…

Quality checking
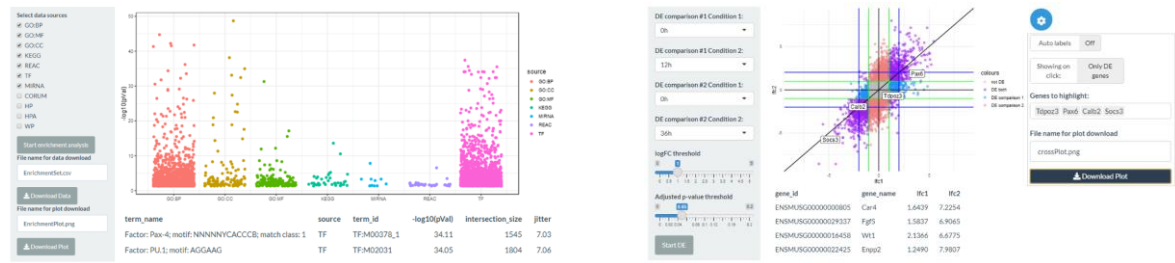
- JSI heatmap

- PCA

Downstream analysis

- Differential expression

- Expression patterns

Functional interpretation

- Enrichment analysis

- Cross plots

However, bulkAnalyseR also features novel gene regulatory network inference

- Gene regulatory networks can be inferred using GENIE3 on the whole dataset or on subsets of conditions to identify changes in topology and strength of regulatory relationship
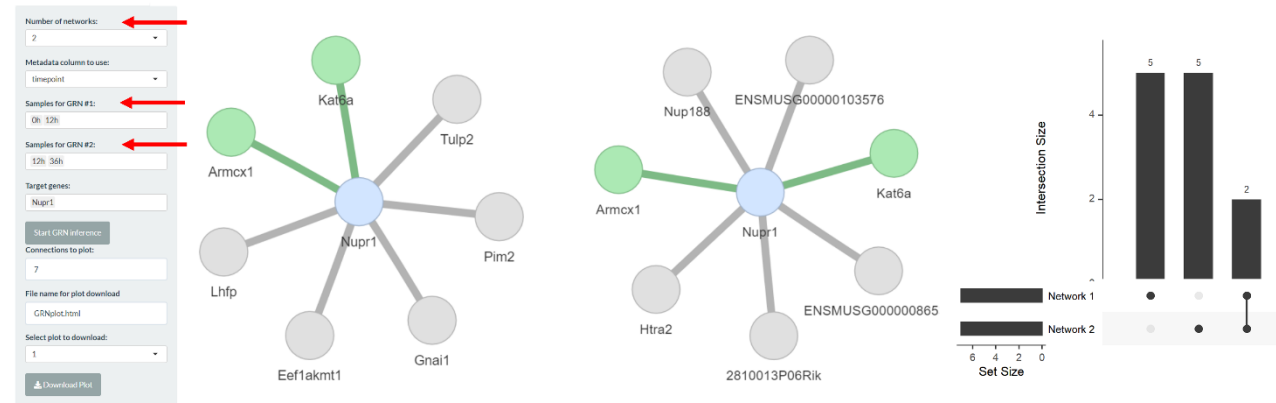
- External information such as GO terms, pathways, transcription factors and miRNAs linking multiple genes in the network can also be used to augment the GRN
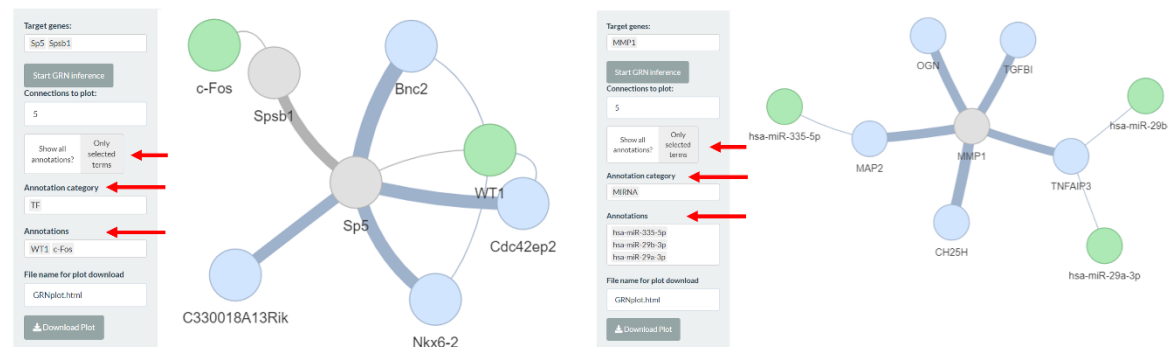
bulkAnalyseR is agnostic the type of expression data provided and can even be used to integrate multiple modalities

- Cis-regulatory relationships can be assessed by connecting entities within a user-defined genomic distances e.g. ChIPseq peaks and differentially expressed genes

- Trans-regulatory relationships are analysed by creating a combined expression matrix across multiple modalities and identifying covariation
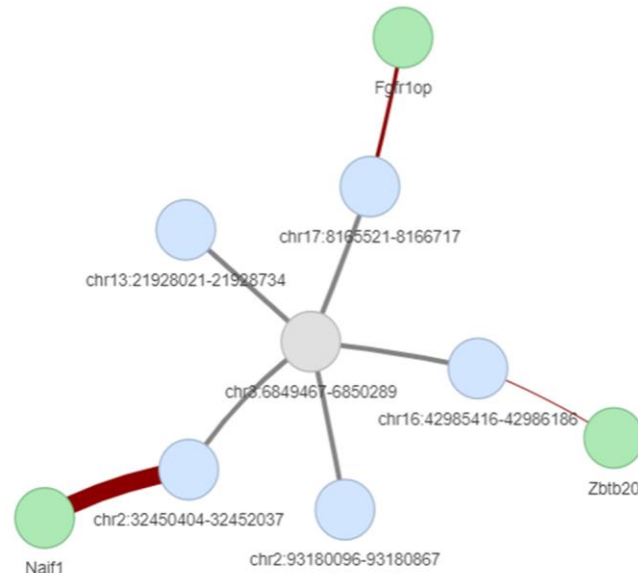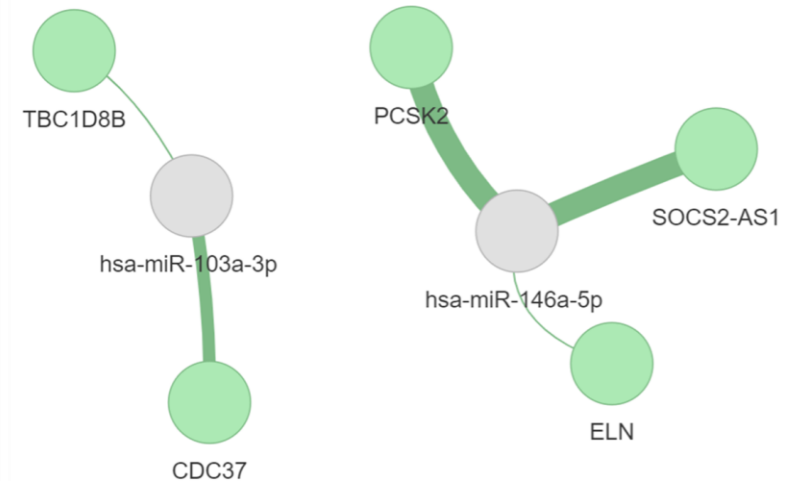
Thanks for listening!

Thanks to my co-authors

Irina Mohorianu

Ilias Moutsopoulos

bulkAnalyseR paper

GitHub

Example app

UNIVERSITY OF CAMBRIDGE