smiRk: New R package for microRNA-mRNA data analysis

Krutik Patel, Carole Proctor, David Young and Daryl Shanley

Newcastle University
k.patel5@newcastle.ac.uk

1. Rationale

Introduction

- MicroRNAs (miRs) control over 60% of mammalian protein coding genes [1].
- Large time course mRNA and miRNA data sets are being generated.
- **I have created an R package to integrate and functionally analyse this type of data.**
- Output of the R package can to help generate hypothesis for further systems or experimental work.

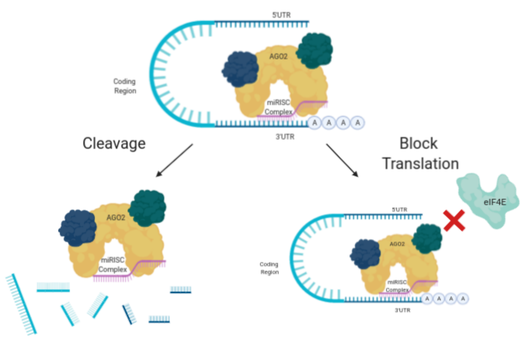

Fig. 1: miRs regulate target mRNA expression levels

Mouse Kidney Fibrosis data [2]

- mRNAseq and small non-coding RNAseq data were located respectively in GSE65267 and GSE61328. Data was processed using standard tools.
- Measurements taken prior to injection of Folic acid and 1, 2, 3, 7 and 14 days after.
- This is a commonly used mouse model for chronic kidney disease experiments.

2. smiRk package pipeline

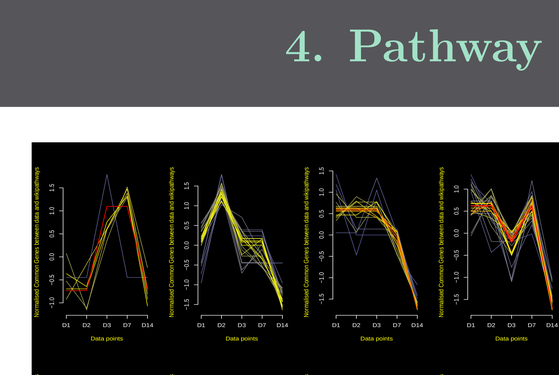

Fig. 2: smiRk package overview

- **Input = differential expression.**
- Wikipathways used for GSE and cluster analysis to identify pathways of interest [6].
- Databases: TargetScans, miRTarBase and miRDB assess potential miR-mRNA interactions [7–9].
- Graphical networks of filtered miR-mRNA interactions can be made in R.
- miR-mRNA interactions found in the pathways of interest can be exported into Pathvisio or Cytoscape [10, 11].
- **Output = Hypothesis generation for further systems or wet lab enquiries.**

3. Time course GSEA

Building upon functions from rWikipathways and clusterProfiler a user can see which pathways their data is most associated with, per time point [3, 4].

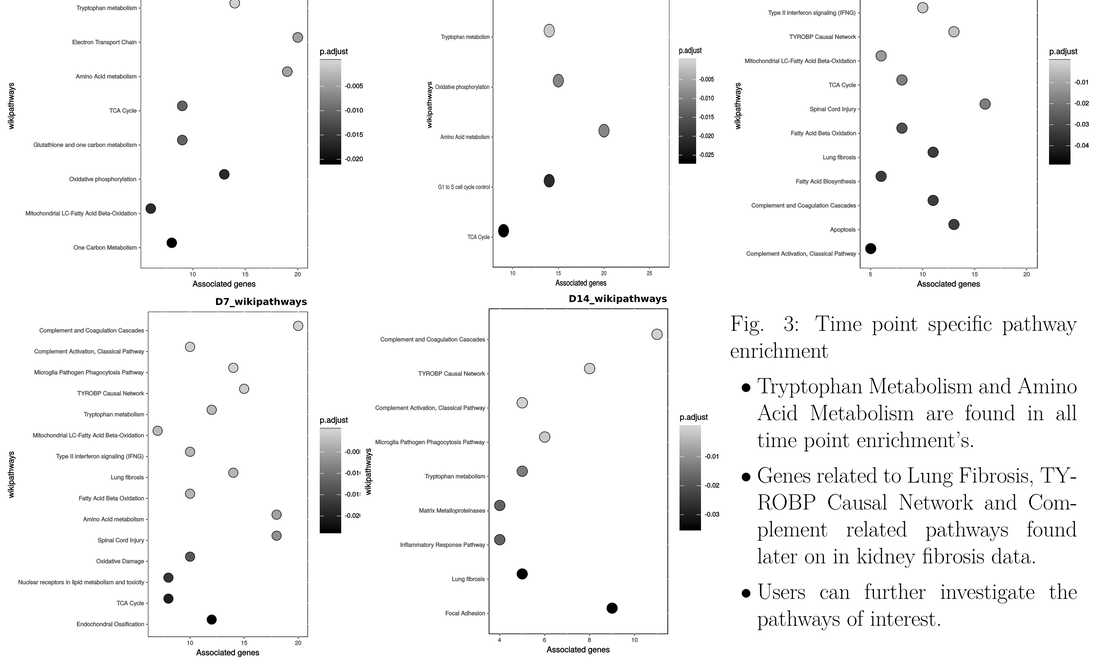

Fig. 3: Time point specific pathway enrichment

- Tryptophan Metabolism and Amino Acid Metabolism are found in all time point enrichment's.
- Genes related to Lung Fibrosis, TYROBP Causal Network and Complement related pathways found later on in kidney fibrosis data.
- Users can further investigate the pathways of interest.

4. Pathway Clustering

Fig. 4: Pathway soft clustering

- Clustering by Mfuzz [12].
- The number of genes found in pathways are clustered by temporal changes.
- Four pathways fitted with cluster 1 at > 0.8 confidence, e.g. Inflammatory Response Pathway.

5. miR-mRNA networks

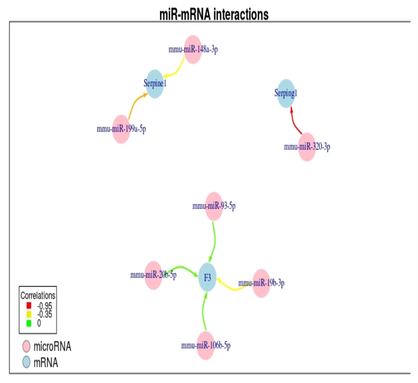

Fig. 5: miR-mRNA interactions in Complement and Coagulation Cascades

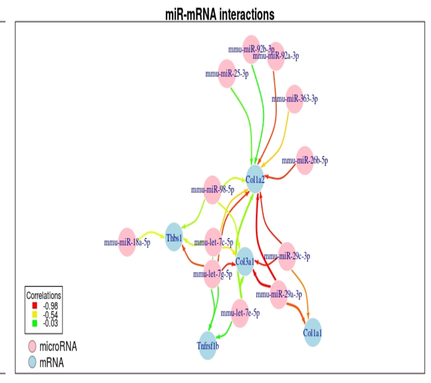

Fig. 6: miR-mRNA interactions in Inflammatory Response Pathway

- Displayed using the igraph package [13].
- User can customise which interactions to view.
- Interactions:
  - negative average correlation
  - In miRDB
  - In TargetScans

6. Pathvisio and Model creation

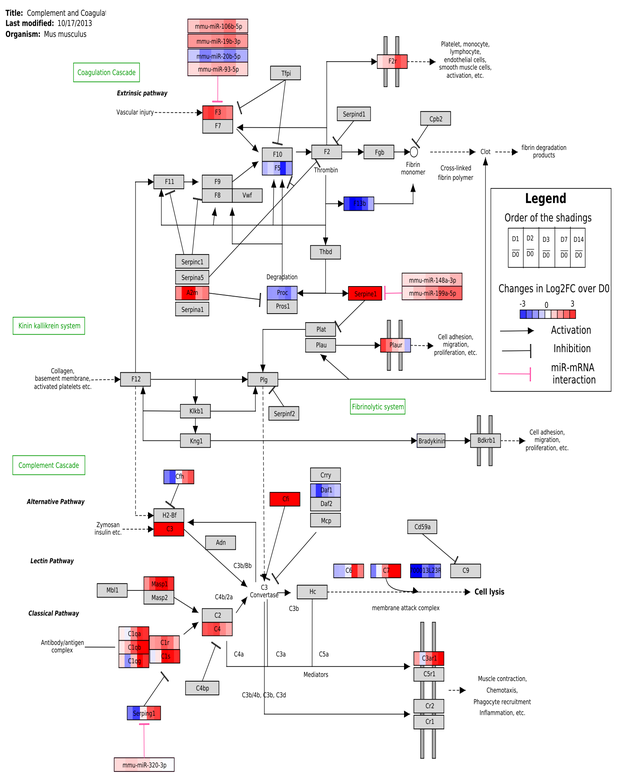

Fig. 7: Pathvisio visualisation of miR integrated Complement and Coagulation Cascades

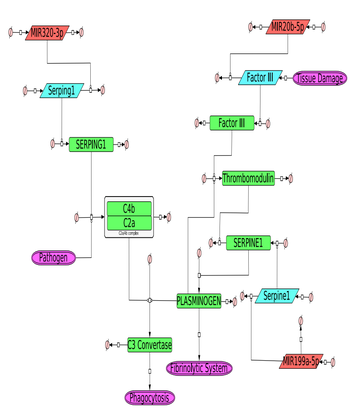

Fig. 8: GRN of a Complement model in Cell Designer

- Pathvisio output can be simplified to build gene regulatory networks (GRNs) on Cell Designer [5].
- smiRk package utilises big multi omic data for model hypothesis generation.
- **Supplements primary use of literature or existing models for model generation.**

Conclusions

Conclusions

- **smiRk package integrates, analyses and generates networks from time series miR-mRNA expression data.**
- **Output can lead to hypothesis generation for systems modelling or experimental work.**
- **smiRk will be accessible from bioconductor in 2020.**

References

[1] Robin C Friedman et al. "Most mammalian mRNAs are conserved targets of microRNAs". In: *Genome research* 19.1 (2009), pp. 92–105.
[2] Mira Pavkovic et al. "Multi omics analysis of fibrotic kidneys in two mouse models". In: *Scientific data* 6.1 (2019), p. 92.
[3] Denise N Slenter et al. "WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research". In: *Nucleic acids research* 46.D1 (2017), pp. D661–D667.
[4] Guangchuang Yu et al. "clusterProfiler: an R package for comparing biological themes among gene clusters". In: *Omics: a journal of integrative biology* 16.5 (2012), pp. 284–287.
[5] Akira Funahashi et al. "CellDesigner 3.5: a versatile modeling tool for biochemical networks". In: *Proceedings of the IEEE* 96.8 (2008), pp. 1254–1265.
[6] Alexander R Pico et al. "WikiPathways: pathway editing for the people". In: *PLoS biology* 6.7 (2008), e184.
[7] Vikram Agarwal et al. "Predicting effective microRNA target sites in mammalian mRNAs". In: *elife* 4 (2015), e05005.
[8] Nathan Wong and Xiaowei Wang. "miRDB: an online resource for microRNA target prediction and functional annotations". In: *Nucleic acids research* 43.D1 (2014), pp. D146–D152.
[9] Chih-Hung Chou et al. "miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions". In: *Nucleic acids research* 46.D1 (2017), pp. D296–D302.
[10] Martijn P van Iersel et al. "Presenting and exploring biological pathways with PathVisio". In: *BMC bioinformatics* 9.1 (2008), p. 399.
[11] Paul Shannon et al. "Cytoscape: a software environment for integrated models of biomolecular interaction networks". In: *Genome research* 13.11 (2003), pp. 2498–2504.
[12] Matthias E Futschik and Lokesh Kumar. "Introduction to Mfuzz package and its graphical user interface". In: (2013).
[13] Gabor Csardi, Tamas Nepusz, et al. "The igraph software package for complex network research". In: *InterJournal, Complex Systems* 1695.5 (2006), pp. 1–9.