Differentially Expressed Genes in Mouse Kidney Cells

A project to identify potential therapeutic targets for fibrosis in human kidney cells

Introduction

- This project makes use of RNA-seq data in order to identify which genes are differentially expressed in mouse kidney cells that present with fibrosis, as opposed to those that do not.
- The mouse is a model organism that is widely used in medical research due to its genetic similarities to humans. Therefore, the differentially expressed genes identified in these cells may have corresponding homologs in humans.
- Therapeutic strategies that target these genes could be further explored in order to treat fibrosis in human kidney cells.

Overview

- *What is RNA-seq analysis?*

RNA-seq analysis is the statistical analysis of data generated from RNA sequencing. It is used to detect the abundance of RNA transcripts in a biological sample.

- *Why is it used?*

RNA-seq has many purposes. This particular project will explore its usefulness in detecting genes with expression levels that are statistically significantly higher in one condition (e.g. a disease) as opposed to another (e.g. the absence of the disease).

- *What is kidney fibrosis?*

Kidney fibrosis is the final manifestation of CKD (chronic kidney disease). It refers to the accumulation of abnormal connective tissue in the kidneys.

- *Is there a genetic basis?*

Research indicates that there may be a genetic component to kidney fibrosis, particularly in inherited kidney diseases that can run in families.

Data Availability and Methods

- The data was obtained from a study of fibrosis in mouse kidney cells and was made available through the ‘RNA-Seq Analysis with Bioconductor in R’ course, offered on the DataCamp platform.

- The data was analysed in R with the `DESeq2` package. After analysing, the Entrez Gene IDs of each gene were converted to their corresponding gene symbols with the `biomaRt` package.

Results

The top 6 most significant genes are presented in the following table:

	geneID	padj	external_gene_name	mgi_symbol
1	18097	0	NA	NA
2	10846	1.59E-259	NA	NA
3	11600	2.03E-164	Angpt1	Angpt1
4	11819	1.20E-146	Nr2f2	Nr2f2
5	30097	1.46E-144	NA	NA
6	26271	1.71E-141	NA	NA

After annotating the genes and converting the Entrez gene IDs to their gene symbols, it was observed that only 2 out of the 6 genes could be identified. As listed, these are the Angpt1 gene and the Nr2f2 gene (genes 3 and 4).

Further Exploration

Known Genes

- Entering the gene IDs of each of the 6 genes into the NCBI search box revealed interesting results.
- Genes 3 and 4, the only 2 annotated genes in the list, were revealed to have RefSeq statuses of REVIEWED and VALIDATED, respectively.
- Moreover, they are known to be protein-coding genes.

Unknown Genes

- Gene 1 was revealed to be Nlf1. Its gene type is currently unknown. In addition, the information was updated as recently as 17 August, 2024.
- This may explain why `biomaRt` was unable to annotate it.
- Genes 4 and 6 were revealed to be ‘predicted’ genes. Gene 6 was revealed to code for snRNA, while the function of Gene 4 is currently unknown.
- This may explain why these two genes could not be annotated.
- The gene ID of gene 5 did not appear to belong to a mouse gene.

Further Exploration

- The amino acid sequences of the proteins encoded by Angpt1 and Nr2f2 were compared with sequences of *Homo sapiens* using the Protein BLAST program and the accession numbers XP_006520386.1 and NP_033827.2 respectively. *Mus musculus* was excluded, while *Homo sapiens* was included in the search. All other parameters were kept at their default values.
- Angpt1 was found to have a homolog in *Homo sapiens* with a percentage identity of 96.76% and an accession number of XP_047277655.1. It is the angiopoietin-1 isoform X1 protein.
- Nr2f2 was found to have a homolog in *Homo sapiens* with a percentage identity of 100% and an accession number of NP_066285.1. This is the COUP transcription factor 2 isoform a protein.
- The results indicate that these homologs are highly evolutionarily similar. They are therefore likely to be structurally and functionally similar.
- The homologous genes could therefore serve as targets for potential gene therapies for kidney fibrosis in humans.

Literature Review

- A review of the existing medical literature appears to support the findings of this analysis.
- Researchers have found evidence that the expression levels Angpt1 and Nr2f2 are elevated in patients presenting with kidney fibrosis.
- One study found that Angpt1 levels were elevated in the plasma of patients with diabetic nephropathy. One of the symptoms of this disease is kidney fibrosis.
- Another showed that deletion of the Nr2f2 gene in mouse models resulted in an improvement in their kidney fibrosis.
- Currently available treatments for kidney fibrosis do not target either of these genes.

Conclusions and Caveats

- The results of the analysis, in combination with research findings, appear to provide evidence that the Angpt1 and Nr2f2 genes may play a role in kidney fibrosis.
- Though these initial findings appear promising, it should be noted that the data was generated from studies of mouse models, which, while generally accurate, are not a substitute for studies on humans.
- Further research is needed on cells from human patients in order to attain conclusive evidence of the therapeutic potential of these genes.

Acknowledgements

- The RNA-seq analysis with the `DESeq2` library was built off of the code provided as part of the course ‘RNA-Seq Analysis with Bioconductor in R’.
- The gene annotation was performed in accordance with the guidelines outlined in the vignette for the `biomaRt` package.

References

Butler, A.E., Al-Qaissi, A., Sathyapalan, T. *et al.* Angiopoietin-1: an early biomarker of diabetic nephropathy?. *J Transl Med* **19**, 427 (2021). https://doi.org/10.1186/s12967-021-03105-9

Wan, R., Long, S., Ma, S. *et al.* NR2F2 alleviates pulmonary fibrosis by inhibition of epithelial cell senescence. *Respir Res* **25**, 154 (2024). https://doi.org/10.1186/s12931-024-02777-3

RNA-Seq with Bioconductor in R

Accessing Ensembl annotation with biomaRt

National Center for Biotechnology Information