Genetic Markers and Pathways in Mice with Kidney Disease

Wenbo Wang, Anil Anderson, Kevin Zhang

Background

- Kidney failure caused by mutations affecting the CoQ enzyme, in turn causes loss of podocyte cells.
- Podocyte cells: Differentiated cells of the kidney glomerulus.
  - Essential cells in kidney filtration
  - Epithelial cells: form covering of all body surfaces, protection, secretion, filtration, excretion.
- Study of kd/kd (kidney disease) mice, similar mutation pathways as humans.
- Modulating the Braf/Mapk pathway with the GDC-0879 compound rescued podocyte injury and kidney filter function.
- GDC-0879 compound
  - Treatment restores an enzyme that protects cells from lipid peroxidation.

Dataset

- From “Targeting a Braf/Mapk pathway rescues podocyte lipid peroxidation in CoQ-deficiency kidney disease” by Sidhom et. al.
- Consists of a set of 6 scRNA-seq data sets corresponding to six 5-month old mice with 3 controls and 3 mice with kidney disease.
- Question of interest: Which gene variants or biological processes are highly expressed in mice genes with kd/kd mutations compared to healthy mice, and if there are implications in human studies for kidney disease?

Imputation

MAGIC (Markov affinity-based graph imputation of cells)

- Method: Multiply UMI count data matrix with exponentiated Markov affinity matrix.
  - Create nearest neighbor graphs (KNN number = 5) with euclidean distance metric.
  - Convert distances to affinities, and a Markov transition matrix M.
  - Exponentiate matrix M by t to represent random walk of length t.
  - Goal: calculate imputed values as M^t times UMI matrix.

SAVER (single-cell analysis via expression recovery)

- Method: UMI count data follow poisson-gamma mixture, or negative binomial model.
  - Poisson: noise
  - Gamma: uncertainty
- α_{gc},β_{gc} are reparameterized as:
  - $\alpha_{gc} = \mu^2/v$, $\beta_{gc} = \mu/v$ from data
- Goal: calculate λ_{gc} (normalized true expression) or posterior probability mean as new UMI count matrix.

$$Y_{gc} \sim Poisson\left(s_c \lambda_{gc}\right)$$
$$\lambda_{gc} \sim Gamma\left(\alpha_{gc}, \beta_{gc}\right)$$

$$\hat{\lambda}_{gc} = \frac{Y_{gc} + \hat{\alpha}_{gc}}{s_c + \hat{\beta}_{gc}}$$

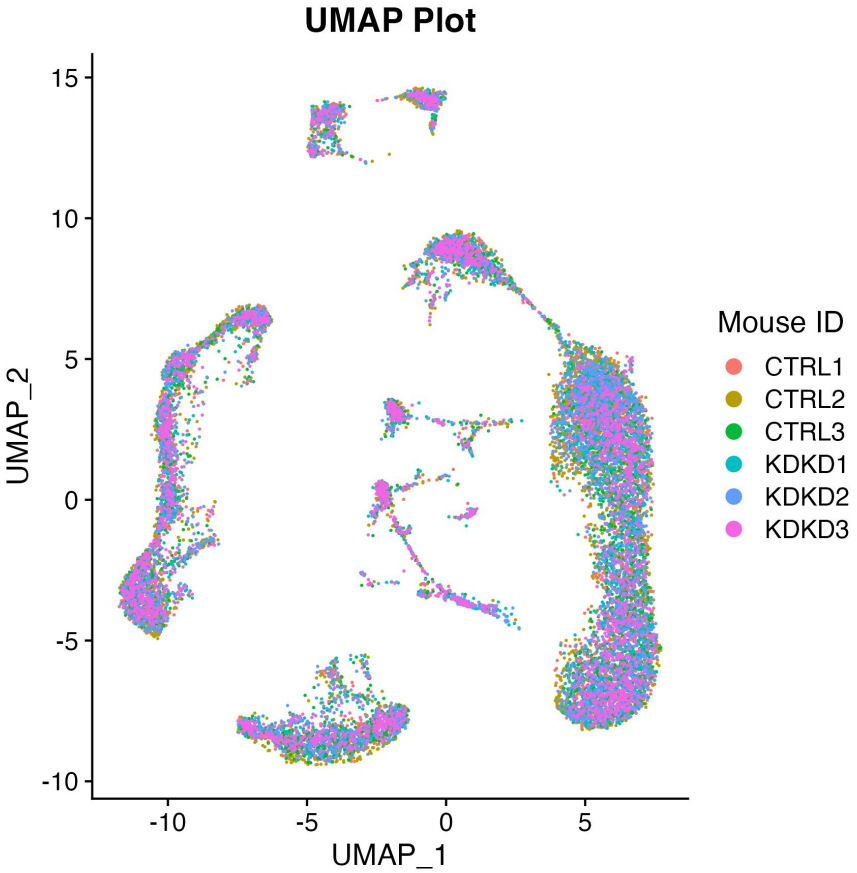
For both methods, 6 datasets are imputed individually. The new UMI count matrices are then processed in Seurat.

Quality Control

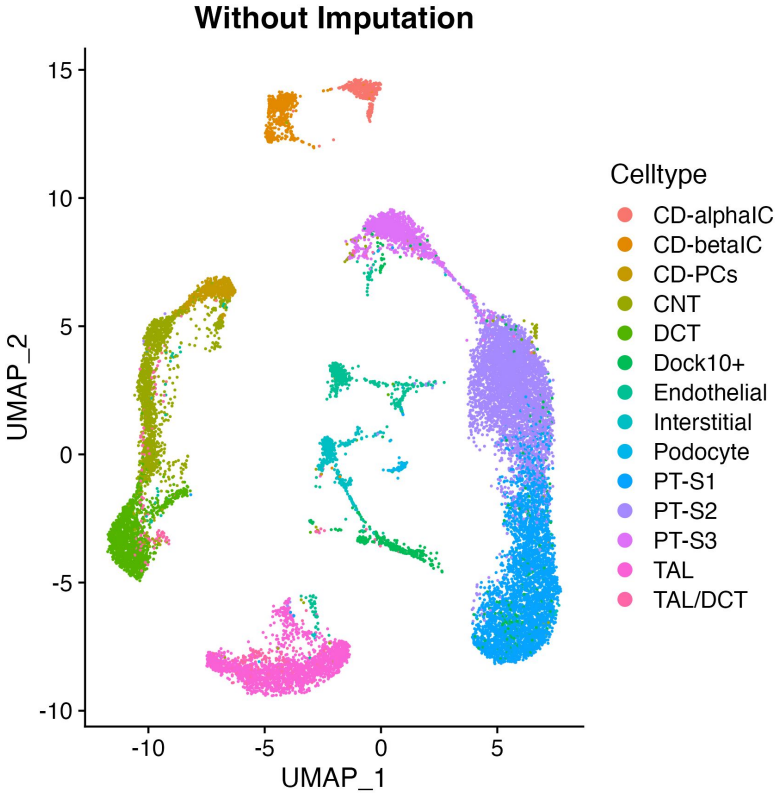
- Took random sample of 3000 cells for each of the 6 mice, but kept all podocyte cells
- Further subsetted using percent mitochondrial genes
- Tried to preserve as much information as possible while also making analysis more computationally feasible

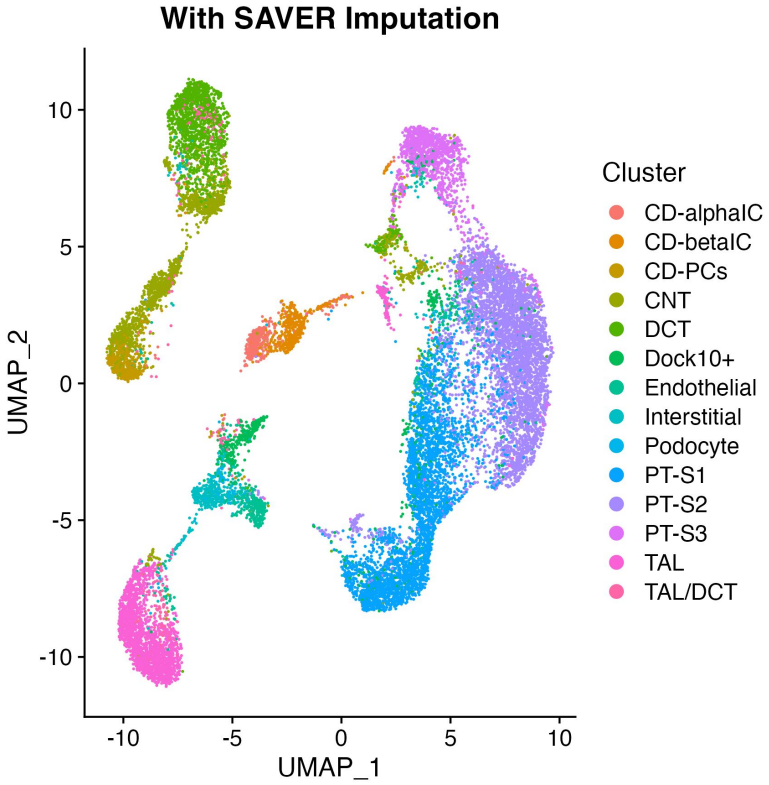
Batch Correction

- Seurat CCA to integrate datasets
- Used integrated assay for dimensionality reduction plots and original RNA assay for differential expression analysis and gene set enrichment analysis



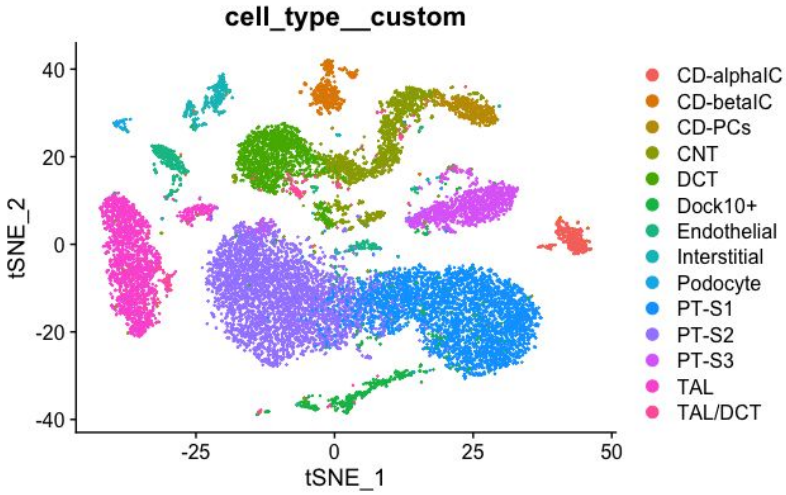
Dimensionality Reduction

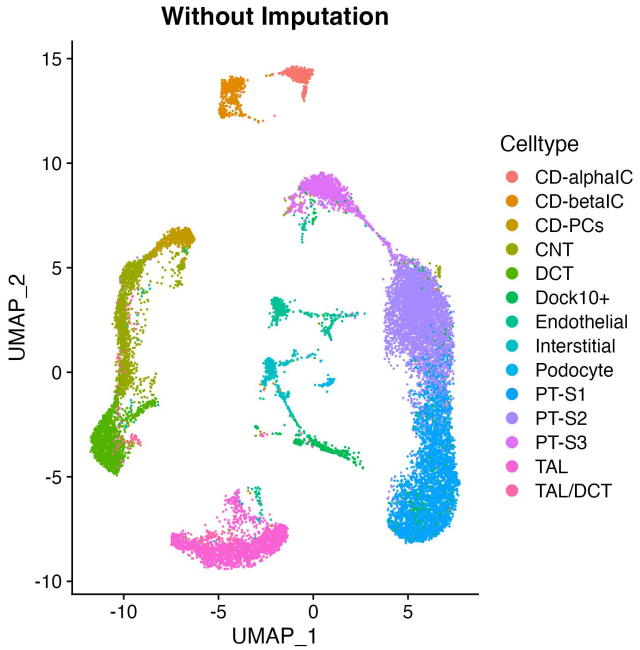




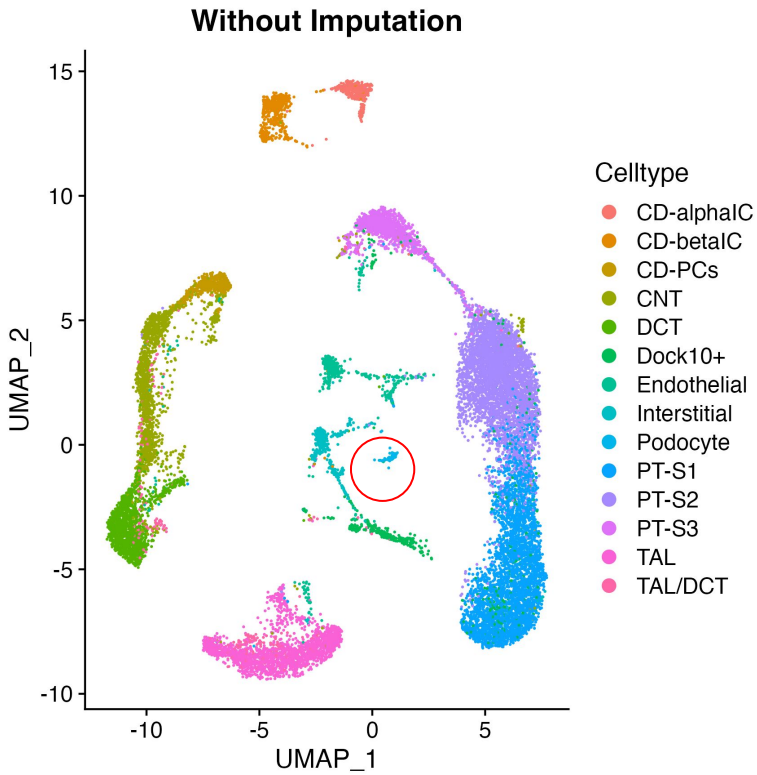
T-SNE Plots

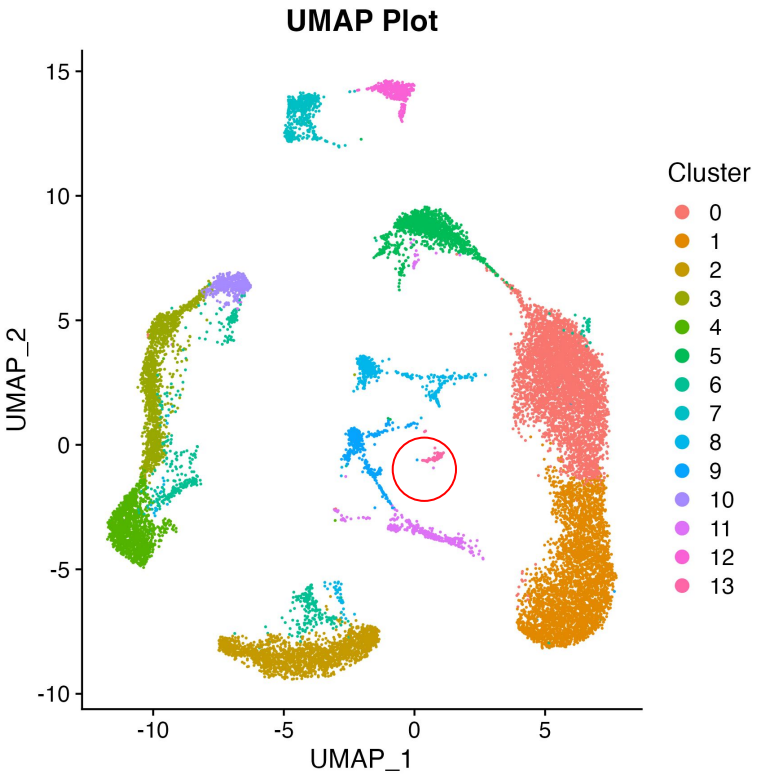
- Separation between clusters relative to UMAP
- Using cell-types



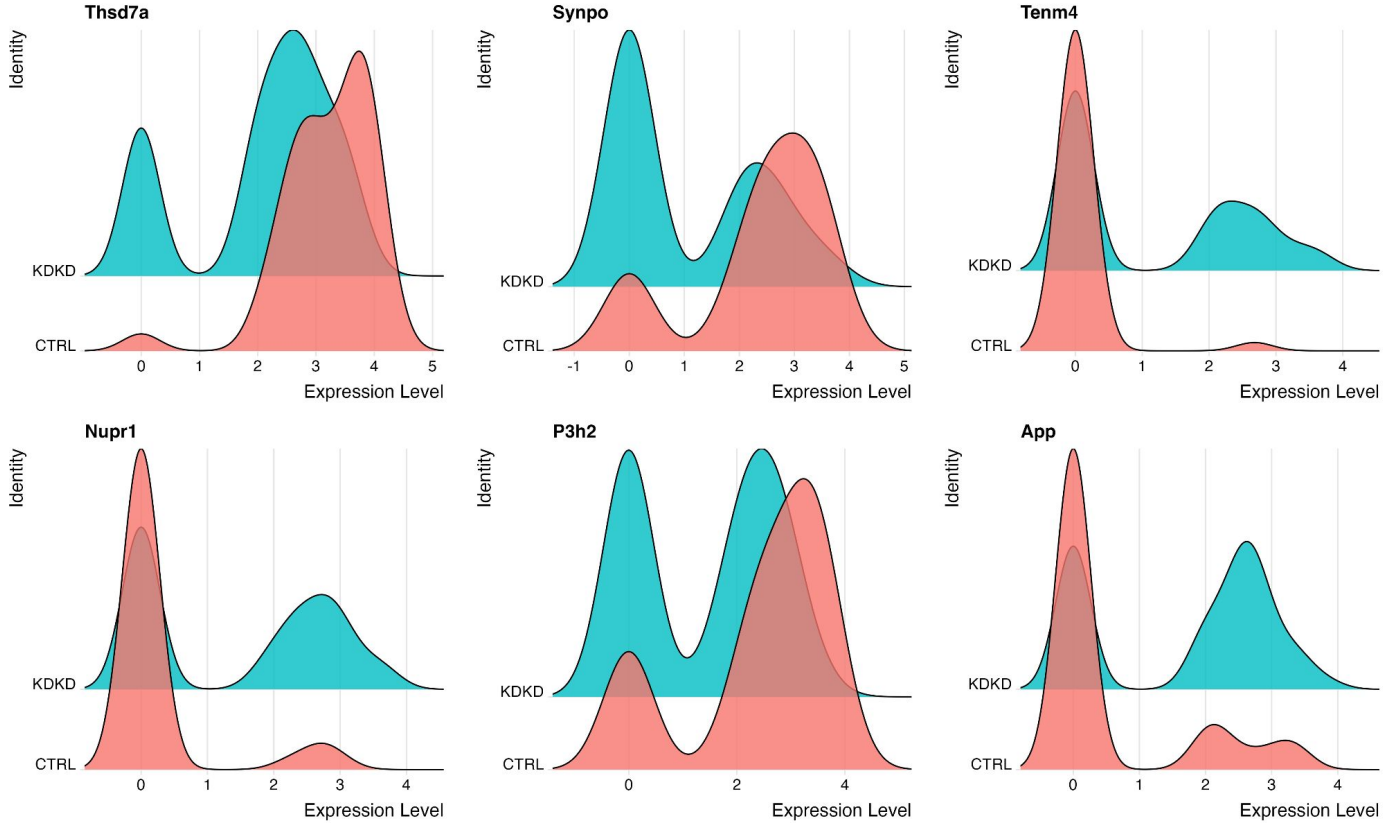


Clustering


Without Imputation
UMAP_2
UMAP_1
15
10
5
0
-5
-10
-10
-5
0
5
Celltype
CD-alphaIC
CD-betaIC
CD-PCs
CNT
DCT
Dock10+
Endothelial
Interstitial
Podocyte
PT-S1
PT-S2
PT-S3
TAL
TAL/DCT

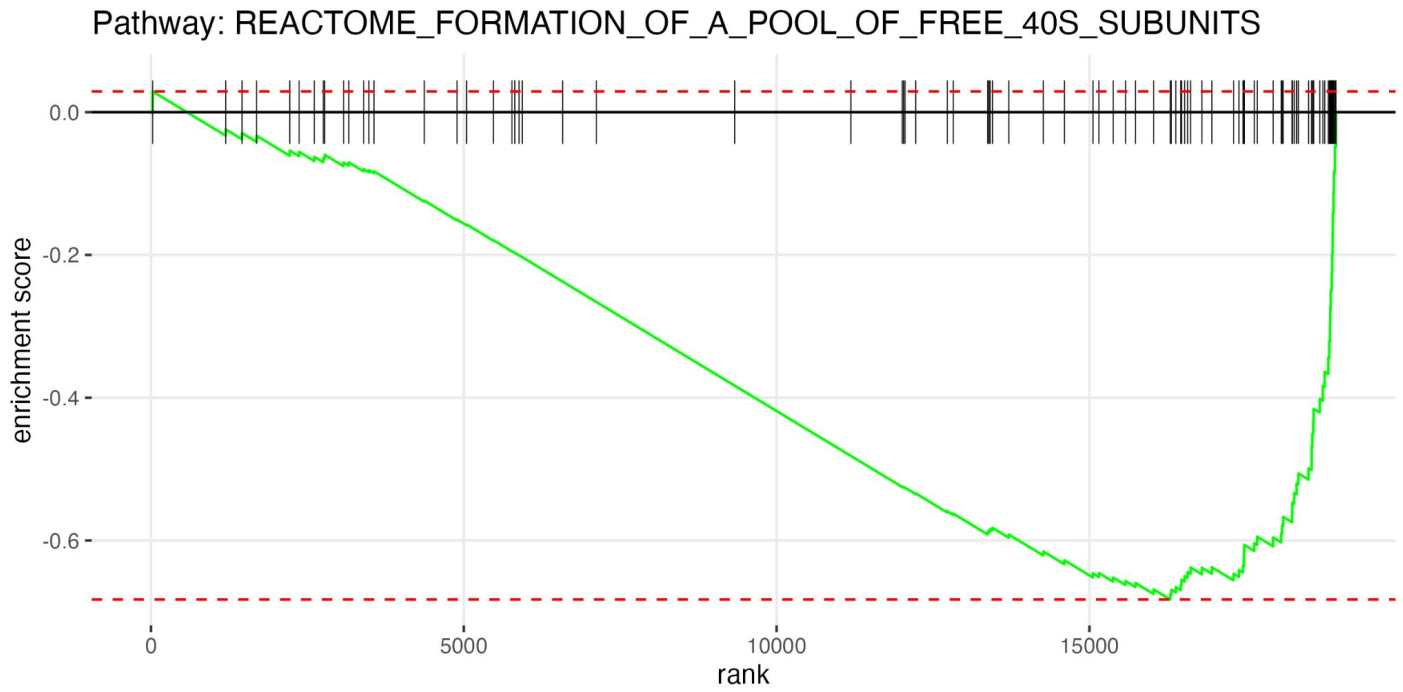

UMAP Plot
UMAP_2
UMAP_1
15
10
5
0
-5
-10
-10
-5
0
5
Cluster
0
1
2
3
4
5
6
7
8
9
10
11
12
13

Podocyte-Specific Differential Expression



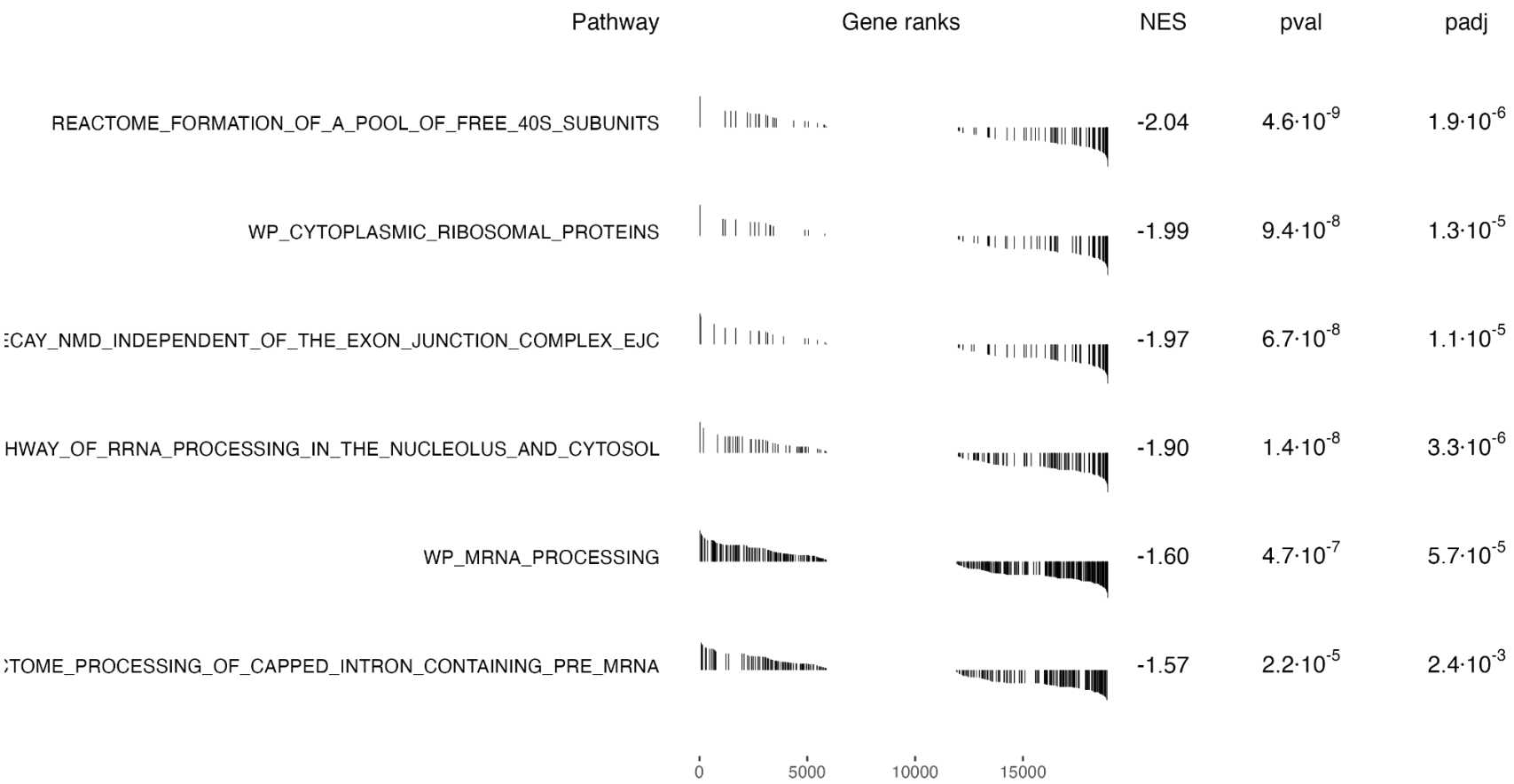
Gene	Regulation
Thsd7a	↓
Synpo	↓
Tenm4	↑
Nupr1	↑
P3h2	↓
App	↑

Gene Set Enrichment Analysis



Leading edge subset: Rps19, Ubb, Rpl38, Rpl32, Rps24, Rps28, Rpl28, Rps8, Rps10, Rpl4, Rpl22, Rplp0, Rpl13a, Rps11, Rps6, Rps29, Rpl23, Rps15, Rpl6, Rpl35a, Rps16, Rps17, Eif2b4, Rpl36a, Rpl31, Rpl18, Rps23, Rpl10, Rpl5, Rpl19, Rps15a, Eif3g, Eif2s3x, Rps14, Rpl8, Eif2s2, Gm2000, Rpsa, Rps9, Rplp1, Eif4g1, Eif3c, Rpl26, Rplp2, Eif3i, Rps3a1, Rps4x, Rpl39, Rpl37, Rpl34, Rpl22l1, Eif5, Rps13, Rpl29

Top Pathways



Discussion

- Our goal was to identify differentially expressed genes in mice with kidney disease vs control mice
- We identified the top marker genes and biological pathways associated with kidney disease
- Top pathway (small ribosomal subunit (40S)) included mostly ribosomal protein genes
    - Is this a true biological effect or simply a result of higher expression of these genes?

Future Work

- Refine imputation and conduct downstream analyses using imputed count matrix.
- Study cell type-specific pathways in non-podocyte cells
- Generalize results to human samples