Supplemental Material (online)

Title: A Single-cell Transcriptome Atlas of the Mouse Glomerulus

Running title: Single-cell transcriptomics of glomerular cells

Authors: Nikos Karaiskos, Mahdieh Rahmatollahi, Anastasiya Boltengagen, Haiyue Liu, Martin Hoehne, Markus Rinschen, Bernhard Schermer, Thomas Benzing, Nikolaus Rajewsky, Christine Kocks, Martin Kann, Roman-Ulrich Müller.

Table of contents - Supplemental Tables and Figures..... page 1

Supplemental Methods ..... page 2

Supplemental Tables and Figures ..... page 6

Supplemental Figure and Table legends ..... page 17

References for Supplemental Material ..... page 20

Supplemental Tables and Figures

Supplemental Figure 1. (Related to Fig. 1)
Quality of glomeruli single-cell data and correlations with bulk mRNAseq data

Supplemental Table 1. (Related to Suppl. Fig. 1B)
Median number of genes, transcripts and reads per cell

Supplemental Table 2. (Related to Suppl. Fig. 1B)
Marker genes for each of the five major cell types to identify putative cell doublets

Supplemental Table 3. (Related to Fig. 1B, Fig. 3A and Fig. 4A)
Top variable genes per cluster

Supplemental Figure 2. (Related to Fig. 2)
Additional markers of glomerular cells

Supplemental Figure 3. (Related to Fig. 3)
Subclustering of glomerular endothelium

Supplemental Table 4. (Related to Fig. 3D and Suppl. Fig. 3E)
Gene sets underlying the gene ontology (GO) term analysis for endothelium subclusters

Supplemental Figure 4. (Related to Fig. 4)
Stress response in podocytes

Supplemental Figure 5. (Related to Fig. 4)
Subclustering of podocytes

Supplemental Figure 6. (Related to Fig. 4)
Pathway and gene set overdispersion analysis for podocyte subclusters

Supplemental Methods

Glomerular isolation and single cell suspension

Isolation of glomeruli and preparation of single-cell suspensions from 8 week old CD1 male mice was performed as previously described[1]. In brief, kidneys were dissected after cervical dislocation and renal arteries were perfused with Hank's Balanced Salt Solution (HBSS) supplemented with magnetic Dynabeads (Thermofisher Sicentific). Next, kidneys were released from the capsule and minced in 3 ml tissue digestion solution containing 300 U/ml Collagenase 4 (Worthington), 300 U/ml Protease and 50 U/ml DNase I. Tissue dissociation was completed in a 37 ºC incubator with gentle shaking and trituration in 5-minute intervals to facilitate the dissociation process. The tissue soup was then passed through 100-um sieve on 50-ml falcon tubes to capture tubular and non-glomeruli structures by their size. The sieves were washed with ice-cold HBSS and the glomeruli were pelleted by spinning down at 1500x g at 4 ºC for 5 minutes. The pellet was resuspended in ice-cold HBSS in 1.5 ml eppendorf tubes and the glomeruli were concentrated using a magnet.
The purity of glomerular isolates was routinely monitored by visual inspection on a Leica DFC450 microscope (Suppl. Figure 1A). For preparation of single cells, isolated mouse glomeruli were re-suspended in 1 ml of tissue digestion solution and transferred to a thermo-shaker set to 37 °C at1400 rpm. Further mechanical dissociation was induced using pipetting and/or gentle shearing through a syringe in 5-minute intervals. After a total time of 40 minutes, beads were removed and the single cells were collected by centrifugation at 5000x g, 4 ºC for 10 minutes. Cells were re-suspended in ice-cold PBS and immediately FACS-sorted. 7-AAD (BD Pharmigen, 559925) was added to stain dead cells. FACS was carried out on a FACSAria III device gating for 7-AAD negative single cells. For fixation, single live cells were collected in ice-cold 80% methanol in PBS and transferred to Berlin on dry ice for the preparation of monodisperse droplets for single cell RNAseq.

Single-cell and bulk mRNA-sequencing

Single-cell sequencing: Drop-seq procedure, single-cell library generation and sequencing

Monodisperse droplets of about 1 nl in size were prepared on a self-built Drop-seq set up following closely the instrument set up and library generation procedures invented by Macosko et al.[2] , as also previously described by us for methanol-fixed cells[3,4]. Upon nanoliter droplet formation, individual cells are co-encapsulated with individual, uniquely barcoded beads, and become lysed completely. Released cellular polyadenylated RNA molecules then hybridize to polyd(T) primers that are attached to the uniquely barcoded beads. Nanoliter-droplets are collected and broken, and RNA molecules are reverse transcribed into complementary DNA (cDNA), amplified by PCR, size-selected for mRNA and sequenced in bulk.
cDNA libraries had large average sizes of ~ 1.7 to 1.9 kb, indicating high quality RNA and cDNA molecules. Single-cell Drop-seq libraries at 1.8 pM (final insert size average ranging between 650 to 780 bp) were sequenced in paired-end mode on Illumina Nextseq 500 sequencers with 1% PhiX spike-in for run quality control using Illumina Nextseq500/550 High Output v2 kits (75 cycles). Read 1: 20 bp (bases 1-12 cell barcode, bases 13-20 UMI; Drop-seq custom primer 1 "Read1CustSeqB"), index read: 8 bp, read 2 (paired end): 64 bp).

Bulk mRNA-seq libraries and sequencing

Isolated glomeruli or live, sorted, methanol-fixed single glomeruli cells were used for total RNA extraction with TRIzol reagent (Invitrogen). Stand-specific cDNA libraries were generated according to the Illumina TruSeq protocol (TruSeq Stranded mRNA LT Sample Prep Kit, Illumina) using 500 ng (glomeruli lysate) or 95 ng (sorted, fixed single cells from dissociated glomeri) of total RNA. 1.8 pM concentrated libraries were sequenced in single-read mode on Illumina Nextseq 500 sequencers using High Output v2 kit (150 cycles), single read 150 bp, index read: 6 bp.

Computational methods

Bulk mRNA-seq: alignment and gene quantification.

Reads were aligned to the Mus musculus reference genome mm10 using STAR v.2.5.3a with default parameters[5]. Gene quantification was done with htseq-count by using the GENCODE M12 (Ensembl 87) annotation. We used the same workflow to align the reads and quantify genes in published datasets (Suppl. Fig. 1F-H).

Single cell RNA-seq: data processing, alignment and gene quantification.

We chose read 1 to be 20bp long, which sequences the cell barcode at positions 1-12 and the UMI at positions 13-20, while avoiding reading into the poly(A) tail. The remaining 64 sequencing cycles were used for read 2. Sequencing quality was assessed by FastQC (v.0.11.2), while special attention was paid to the base qualities of read 1 to assure accurate cell and UMI calling. We used the Drop-seq tools v. 1.12[2] to trim poly(A) stretches and potential SMART adapter contaminants from read 2, to add the cell and molecular barcodes to the sequences and to filter out barcodes with low quality bases. The reads were then aligned to the reference genome as above. Typically, around 60%-75% of the reads were found to uniquely map to the genome; muti-mapping reads were discarded. The Drop-seq toolkit[2] was further used to add gene annotation tags to the aligned reads and to identify and correct bead synthesis errors, in particular base missing cases in the cell barcode. Cell numbers were estimated by plotting the cumulative fraction of reads per cell against the cell barcodes and calculating the knee point. The DigitalExpression tool[2] was used to obtain the digital gene expression matrix (DGE) for each sample.

Cell filtering and data normalization.

For a first view of gene quantification and basic statistics we used the R package 'dropbead' (https://github.com/rajewsky-lab/dropbead)[3]. The individual DGEs corresponding to four replicate samples were pooled together, resulting in 15,135 cells. We discarded cells expressing less than 350 UMIs, ending up with 14,722 for further analysis. We normalized the UMI counts for every gene per cell by dividing its UMI count by the sum of total UMIs in that cell, and multiplying it by the number of UMIs that the deepest cell contained. Downstream analysis was performed in log space.

Doublet identification and removal.

We used the marker genes in Table 1 and the following algorithm to identify and remove cell doublets[4]. First, for each of the marker genes, we computed its mean expression only from cells which express it. Second, for a given cell, we used these computed averaged values to scale the expression of any of the marker genes expressed in that cell, producing scaled scores. This has the effect of not favoring outlier, or very highly expressed genes. Next, we computed the mean score of every cell type for every cell. If a cell had two or more cell type scores larger than a cutoff, it was flagged as a doublet. As we expected around ~10% of doublets based on earlier cell mixing experiments with distinct species[3,4], we took a threshold value of 0.08, identifying 1,754 doublets in our data. We removed those cells to continue our analysis.

Correlation of gene expression measurements.

Correlations of gene expression levels between single-cell samples were computed by first sub-setting the DGEs of the two samples to the intersection of the genes captured in both libraries (typically more than ~15,000) and then computing the sum of gene counts across all cells in each library. Plotting of correlations is shown in log-space. For the correlation of our

single-cell libraries against mRNA-seq, we converted the gene counts of the latter one into RPKMs and used the mean value of all isoforms for a given gene.

Clustering, t-SNE representation and marker discovery.
We used Seurat[6] v. 2.2.0 to identify highly variable genes, perform principal component analysis, identify the most important principal components and use those for clustering and t-SNE representation. Around 1305 highly variable genes were identified and the first 15 principal components were used for the clustering and the tSNE representation. Cluster identification was done with the Louvain clustering algorithm. The endothelial and podocyte subclusters were found to be relatively robust to changes of the resolution parameter. We used Seurat's function FindAllMarkers to find potential markers for each cluster in an unsupervised manner.

Detection of tissue dissociation-induced stress response
A set of 140 genes, recently shown to respond to tissue dissociation stress, was taken from van den Brink et al. 2017[7] (Suppl. Table 5). In order to assess whether our single-cell dissociation induced a stress response, we compared our two bulk mRNA-seq samples ("bulk1" corresponding to purified glomeruli; "bulk2" corresponding to sorted, fixed single cells). The difference of normalized gene expression between the two samples was used to measure the stress response. We restricted the analysis to those genes having a normalized expression greater than 1 in both bulk samples, resulting in 132 out of 140 initial genes. We created a second set by randomly selecting an equal number of genes with the same requirement. A third set of genes was created by excluding the stress response genes from the highly variable genes and requiring again a normalized expression greater than 1 (975 genes in total). Normality of the three distributions was assessed with a Kolmogorov-Smirnov test ($p<2.2x10^{-16}$, $p=1.9x10^{-4}$, $p<2.2x10^{-16}$). The differences in means between the three sets was found to be statistically significant, if tested with a t-test ($p<2.2x10^{-16}$ for each comparison).

Subclustering and correction for tissue dissociation-induced stress response genes
We subclustered the podocytes and the endothelial cells by extracting them from the main clustering and treating them as independent samples. The corresponding raw DGEs (with UMI counts) were used as input to Seurat. As the podocytes subclusters were defined by subtle differences in gene expression, we regressed out the stress response genes to alleviate any stress-related confounding factors. We found that the removal of those genes from the highly variable genes of the Seurat object was the most effective way to regress out the stress factor.

Functional annotation by gene ontology analysis
We used pathway and gene set overdisperson analysis (PAGODA)[8] to detect functional characteristics across cell populations in endothelial and podocyte subclusters. The color bars above PAGODA heatmaps (Fig. 3C, Suppl. Fig. 4H) were defined based on cell cluster colors shown in the tsne plots (Fig. 3A, 4A). Both predefined KEGG pathways and de novo gene clusters were tested. For significant over-dispersed de novo gene clusters, gene ontology analysis was performed using clusterProfiler[9] with pAdjustMethod="BH", pvalueCutoff=0.01, and qvalue = 0.05.

Image analysis and microscopy

Literature and Human Protein Atlas review for identification of novel marker genes
Cell cluster marker genes as identified by highly variable gene analysis between clusters were classified in three categories: 'established marker', 'relevant as by published literature', and 'novel'. The search strategy was designed to identify a comprehensive list of novel cell-type specific markers for the three key glomerular cell types. Ideally, such novel markers would be corroborated by immunostaining data from the Human Protein Atlas (HPA)[10].

In a first step, disease genes causing hereditary kidney disease upon mutation and genes widely used as markers for the respective cell population in immunohistochemistry and immunofluorescence staining were classified as 'established marker'. Then, violin plots for marker expression in glomerular cells were visually inspected to identify marker genes showing exclusive expression in either podocytes, mesangial, or endothelial cells. Genes that either did not have corresponding data in HPA, showed ubiquitous expression in HPA, or were classified as 'not detected' or 'low expression' in glomeruli within HPA, respectively, were not followed up further.

The remaining genes were searched for in the Pubmed database as current in November 2017 using "[*(gene symbol)* OR *(gene name)* OR *(protein name)*] AND endothel* AND (kidney OR glomerulus)", "[*(gene symbol)* OR *(gene name)* OR *(protein name)*] AND podocyt* AND (kidney OR glomerulus)", and "[*(gene symbol)* OR *(gene name)* OR *(protein name)*] AND mesangi* AND (kidney OR glomerulus)", respectively, as search terms. Upon identification of at least one abstract relating to a function of the gene in question in glomeruli or kidneys *in vivo*, the respective gene was classified as 'relevant as by published literature'. If no publications were identified in this manner Pubmed was searched again using "[*(gene symbol)* OR *(gene name)* OR *(protein name)*] AND (kidney OR glomerulus)" as search term. If no publications were found or the putative marker gene had thus far only been shown as relevant *in vitro*, the marker was classified as 'novel'.

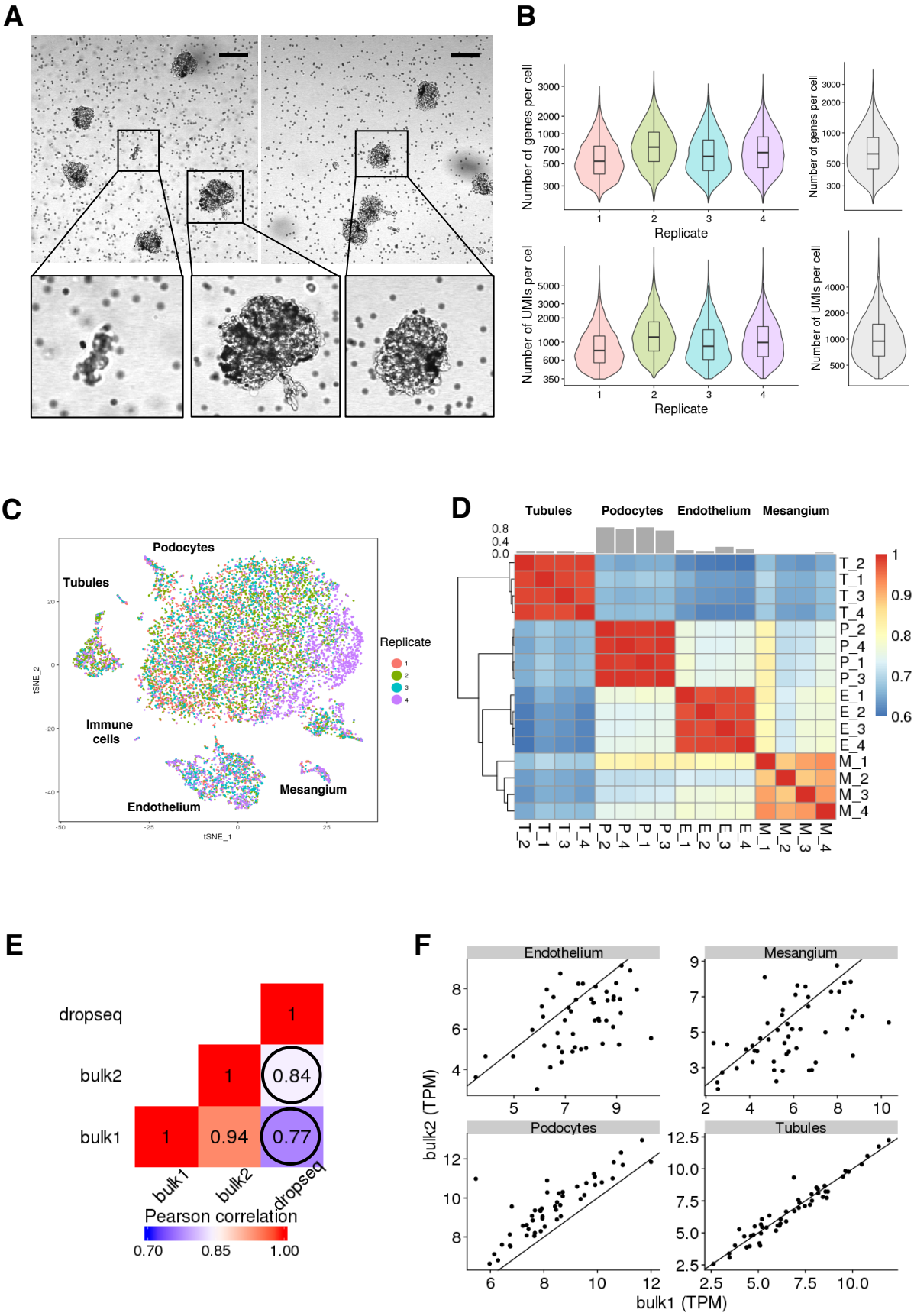
Immunofluorescence staining & image processing

Immunofluorescent staining of isolated glomeruli was carried out on 8 week old male Nphs2-Cre x mTmG mice in a CD1 background[11]. Fixation was performed with 10% PFA followed by permeabilization with Triton-based PBS at room temperature for 30 minutes. Primary antibody incubation was completed overnight at 4 ºC with Cald1 antibody (Sigma, HPA008066), Lars2 antibody (Abcam, ab187983) or an isotype control (Santa Cruz, sc-2027).

Confocal stacks were acquired using a Leica SP8 confocal microscope equipped with a 63x/1.4 oil immersion objective operated with the LAS X software version 3.1.5 (Leica Microsystems, Wetzlar, Germany). Image acquisition settings were comparable between all images.

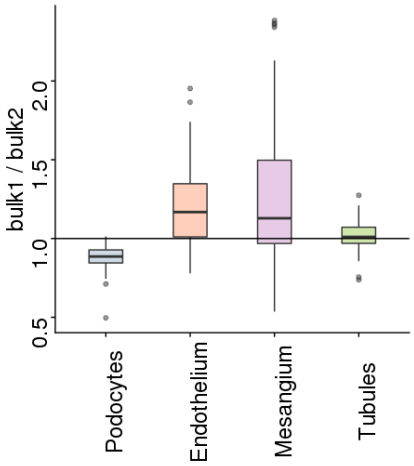
To prepare representative figures images were processed using ImageJ/Fiji (version 1.51) as follows: Maximum intensity projections of 5 slices spanning 1.49µm. Display settings (Min&Max adjustement) in the control channels (mTomato, mGFP, DAPI) were individually adjusted for best result, while the settings were kept identical for the channel of interest (Cald1, Lars2 and control). Montages were prepared using the ImageJ/Fiji Montage tool.

Supplemental Figure 1

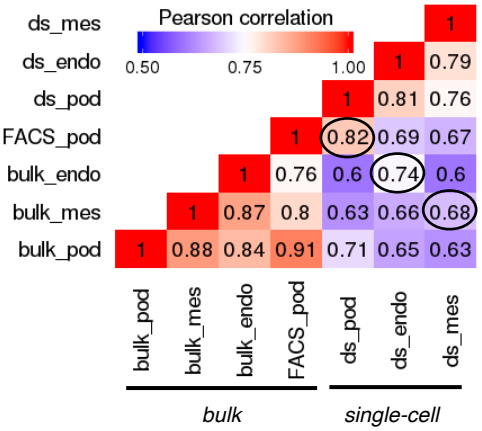


Supplemental Figure 1

G



H



Supplemental Table 1. Median number of genes, transcripts (UMIs) and reads per cell

		reads	cells	genes	umis
before	**replicate1**	8077	4394	532	786
doublet	**replicate2**	10091	4164	743	1156
removal	**replicate3**	8982	2948	595	884
	replicate4	10562	3216	649	986
mean		9428	3681	630	953
sum			14722		
after	**replicate1**	8294	3815	544	806
doublet	**replicate2**	9744	3768	730	1127
removal	**replicate3**	8963	2553	589	878
	replicate4	10496	2818	648	983
mean		9374	3239	628	949
sum			12954		

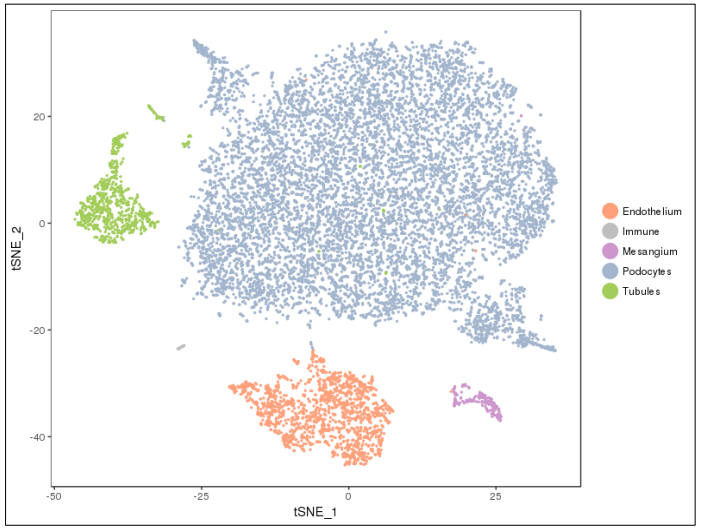
Supplemental Table 2. (Related to Suppl. Fig. 1B). Marker genes for each of the five major cell types to identify putative cell doublets.

Endothelium	Mesangium	Podocytes	Tubules	Immune
Emcn	*Acta2*	*Nphs2*	*Fxyd2*	*Cd74*
Ehd3	*Rgs5*	*Cdkn1c*	*Pvalb*	*H2-Aa*
Kdr	*Myl9*	*Tcf21*	*Atp1b1*	*Cd52*
Flt1	*Tpm2*	*Enpep*	*Defb1*	*Ptprc*
Pecam1	*Myh11*	*Nphs1*	*Umod*	
Pi16	*Pcp4l1*	*Synpo*	*Wfdc15b*	
S100a6	*Sfrp2*	*Npnt*	*Slc12a3*	
Pbx1	*Rergl*	*Wt1*	*Sfrp1*	
	Pln	*Pard3b*	*Ldhb*	
	Ren1	*Ptpro*	*Hsd11b2*	
		Iqgap2	*Kcnj1*	
		Mafb		

Supplemental Table 3. Top variable genes per cluster

Full dataset

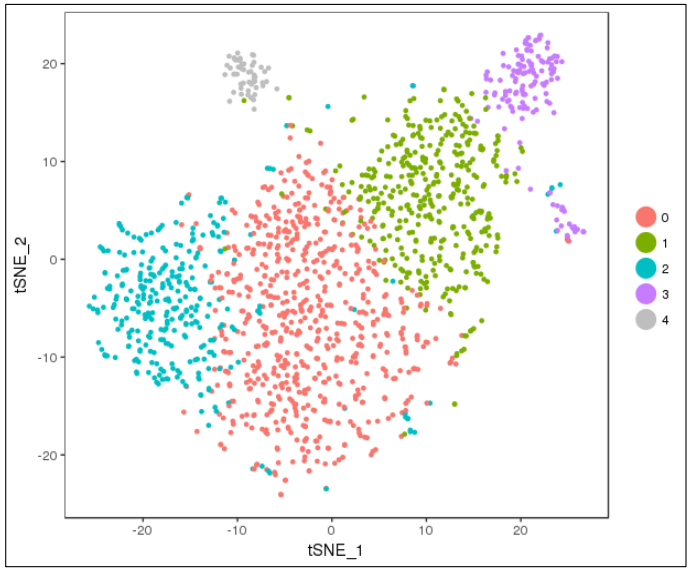
Endothelium.avg_logFC	Endothelium.gene	Immune.avg_logFC	Immune.gene	Mesangium.avg_logFC	Mesangium.gene	Podocytes.avg_logFC	Podocytes.gene	Tubules.avg_logFC	Tubules.gene
3.18	Ly6c1	4.1	Cd74	3.75	Acta2	2.53	Nphs2	4.02	Fxyd2
3.07	Srgn	4.1	H2-Aa	3.43	Myl9	2.48	Cdkn1c	3.96	Klk1
2.87	Emcn	3.9	H2-Ab1	3.40	Rgs5	2.16	Clic3	3.89	Defb1
2.77	Egfl7	3.5	H2-Eb1	3.38	Tagln	2.14	Nupr1	3.85	Atp1b1
2.74	Cdkn1a	3.0	Lyz2	3.09	Tpm2	1.99	Dpp4	3.82	S100g
2.71	Ifitm3	2.6	Tyrobp	2.96	Myh11	1.98	Enpep	3.45	Pvalb
2.70	Ednrb	2.5	Lilrb4a	2.93	Sparcl1	1.93	Tcf21	3.28	Spp1
2.60	Apold1	2.5	Laptm5	2.73	Mustn1	1.91	Nphs1	3.24	Umod
2.58	Igfbp5	2.4	Ccl9	2.72	Gm13889	1.87	Gadd45a	3.23	Slc12a3
2.57	Ctla2a	2.3	Fcer1g	2.53	Serpine2	1.86	Rab3b	3.14	Mt1
2.49	Cd24a	2.2	Cd52	2.45	Hopx	1.76	Rhpn1	3.10	Ldhb
2.39	Emp1	2.0	Ctss	2.38	Akr1b7	1.75	Tmsb4x	3.08	Wfdc15b
2.31	Ly6e	1.8	Ptprc	2.27	Pcp4l1	1.75	Col4a3	3.06	Sfrp1
2.24	S100a6	1.8	Cd53	2.18	Map3k7cl	1.74	Rasl11a	3.05	Calb1
2.23	Adgrf5	1.9	Cd83	2.10	Rergl	1.74	Mafb	2.87	Mt2
2.19	Flt1	2.3	Cytip	1.83	Pln	1.72	Npnt	2.70	Nudt4
2.16	Ehd3	1.4	Lsp1	1.82	Fxyd1	1.70	Arhgap24	2.66	Tmem213
2.12	Clic4	2.0	Srgn	2.16	Rasd1	1.70	Adm	2.55	Ppp1r1a
2.11	Crip1	1.9	Lgals3	2.15	Mgp	1.69	Pak1	2.55	Clu
2.06	1810011O10Rik	2.1	Rgs2	1.88	Sncg	1.68	Synpo	2.50	Tmem52b
2.05	Gimap6	1.7	Gm2a	1.67	Mef2c	1.65	Foxd2os	2.36	Kng2
2.04	Plpp1	1.5	Ly6e	1.64	Flna	1.63	Golim4	2.34	Egf
2.02	Slfn5	1.3	Fxyd5	2.13	Id3	1.61	Igfbp7	2.29	Wfdc2
1.94	Fxyd5	1.5	Ctsz	1.86	Gng11	1.60	Vegfa	2.22	Atp1a1
1.94	Pecam1	1.4	Gm8995	2.49	Cald1	1.59	Cd59a	2.17	mt-Cytb
1.94	Gm8995	1.3	Rpl18a	1.87	Dkk2	1.59	Sdc4	2.15	Gabarapl1
1.94	Pi16	1.4	Cyba	1.86	Nr4a2	1.58	Sema3g	2.14	Fabp3
1.93	Gimap4	1.4	Arpc1b	1.87	Rgs2	1.57	Tdrd5	2.09	Kl
1.92	Sgk1	1.5	B2m	1.79	Slc12a2	1.56	Nap1l1	2.04	Mal
1.92	Kdr	1.4	Gpx1	1.88	Filip1l	1.54	Shisa3	2.04	Kcnj1
1.90	Ptprb	1.3	Gm15427	1.47	Usp2	1.54	Eif3m	2.01	Pgam2
1.89	Ramp2	1.3	Emp3	1.77	Crip1	1.48	Thsd7a	1.98	mt-Nd1
1.87	Klf4	1.4	Rpl32-ps	1.43	Akap12	1.48	Pth1r	1.97	mt-Co1
1.86	Pbx1	1.4	H2afz	1.52	Id2	1.48	Sept11	1.90	S100a1
1.86	Thbd	1.3	H2-K1	1.50	Mt1	1.47	Ctsl	1.90	Oxct1
1.85	Cyp4b1	1.3	Rpsa	1.07	Cdkn1a	1.46	Podxl	1.89	Chchd10
1.84	B2m	1.3	Rps4x	1.78	Impdh2	1.45	Cryab	1.87	Wnk1
1.83	Calcrl	1.2	Gm10689	1.32	Adamts1	1.44	Mertk	1.87	mt-Nd4
1.78	Ly6a	1.3	Gm29228	1.39	Igfbp5	1.44	Htra1	1.84	Slc25a5
1.78	Meis2	1.6	Rilpl2	1.50	Ppp1r12a	1.42	Nes	1.84	Sostdc1
1.78	Hbegf	1.2	Rps13-ps2	1.20	Tpm1	1.52	Wt1	1.83	mt-Nd2
1.76	Lrrc32	1.2	Gm9354	1.38	Actn1	1.55	Npr3	1.83	Epcam
1.76	Crip2	1.6	Msn	1.41	Emd	1.49	Ildr2	1.80	Clcnkb
1.70	Sdpr	1.3	Rpl23	1.19	Mylk	1.52	Robo2	1.80	Slc16a7
1.63	Clic1	1.2	Rpl32	1.08	Slc25a4	1.54	Pard3b	1.78	mt-Nd5
1.54	Nrp1	1.3	Gm9794	1.10	Gadd45b	1.48	Tmem150c	1.77	mt-Rnr1
1.48	Gpx1	1.2	Gm6472	1.13	Dstn	1.54	Gas1	1.77	Acat1
1.51	Plpp3	1.3	Rap1b	1.14	Cebpb	1.46	Hoxc8	1.76	Gm5514
1.40	Akap13	1.2	Rps24	1.07	Map1lc3a	1.50	Iqgap2	1.75	Pdzk1ip1
1.37	Mat2a	1.3	Cst3	1.13	Cstb	1.44	Sema3e	1.74	Ly6a



Supplemental Table 3. Top variable genes per cluster

Endothelium subclusters

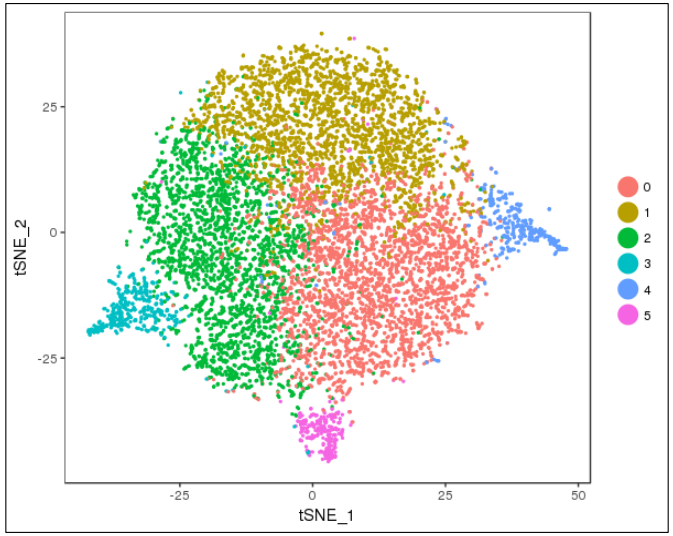
0.avg_logFC	0.gene	1.avg_logFC	1.gene	2.avg_logFC	2.gene	3.avg_logFC	3.gene	4.avg_logFC	4.gene
								residual doublets	
0.85	Plat	2.09	Calca	0.98	Klf2	3.10	Mgp	1.96	Tcf21
0.83	Ednrb	1.98	Tspan8	1.20	Hspa1a	2.37	Fbln5	2.26	Clic3
0.96	Ctla2a	1.61	S100a4	1.23	Hspa1b	2.22	Jag1	2.63	Nphs2
0.71	Egfl7	0.83	Crip1	0.85	Jun	1.81	Igf2	1.78	Gsta4
0.51	Cd24a	0.89	Tm4sf1	0.99	Hsp90aa1	1.82	Cldn5	1.79	Ptpro
1.12	Akap12	0.99	Vim	1.34	Hsph1	1.87	Cxcl1	1.72	Rab3b
0.49	Igfbp5	1.24	Sox17	0.84	Fos	1.50	Egr1	1.78	Myom2
0.64	Marcksl1	0.68	S100a10	0.96	Sgk1	1.60	Fbln2	2.21	Gas1
0.50	Mat2a	1.08	Plac8	0.78	Egr1	1.31	Vim	2.14	Enpep
0.59	Litaf	0.76	S100a6	0.88	Ehd3	1.26	Klf2	1.87	Dpp4
0.58	Kdr	0.63	Ly6a	0.70	Fosb	1.28	Jun	1.53	Pdpn
0.31	Emcn	0.91	Palmd	1.02	Dnajb1	1.30	Fos	2.47	Igfbp7
0.61	S1pr1	0.77	Thbd	0.95	Ppp1r15a	0.83	Tmsb4x	1.53	Foxd2os
0.39	Ehd3	0.69	Slc6a6	0.93	Id3	1.18	Cd93	1.83	Gadd45a
0.58	Cyp4b1	0.75	Glul	0.69	Junb	1.15	Zfp36	2.53	Cdkn1c
0.37	Cdkn1a	0.66	Calcrl	0.91	Gm12346	1.19	Ier2	2.11	Sdc4
0.42	Tpm3	0.56	Fxyd5	0.92	Socs3	1.24	Id3	1.67	Tdrd5
0.45	Uqcrq	0.53	Anxa2	0.85	Zfp36	1.21	Stmn2	1.61	Pth1r
0.55	Lrrc32	0.39	Ifitm3	0.53	Hsp90ab1	1.21	Cyr61	1.84	Arhgap24
0.30	Lmna	0.57	Ahnak	0.98	Rhob	1.22	Slco2a1	1.73	Shisa3
0.33	Plpp1	0.52	Arpc1b	1.12	Plk2	0.84	Ly6a	1.56	Wt1
0.58	Cdk11b	0.60	Tinagl1	0.70	Hspa8	1.30	Fam3c	1.55	Itgb5
0.26	Srgn	0.50	Akap13	0.47	Klf4	1.04	Aggf1	1.65	Nphs1
0.42	Plpp3	0.53	S100a16	0.60	Igfbp5	0.99	Id2	1.41	Rhpn1
0.43	Tpm3-rs7	0.51	Uaca	0.50	Atf3	0.96	Utrn	2.12	Nupr1
0.32	Nrp1	0.39	Syn3	0.56	Dnaja1	1.06	Ptrf	1.47	Cryab
0.46	Hes1	0.45	Adgrl4	0.52	Hspb1	1.01	Sema3g	1.54	Mafb
0.31	Hbegf	0.35	Anxa1	0.69	Btg2	0.95	Nr4a1	1.55	Ctgf
0.35	Ppp1r2	0.44	Ifi203	0.69	Sertad1	0.76	Atf3	1.51	Ezr
0.30	Srsf2	0.40	1810011O10Rik	0.47	Plpp1	0.56	Ly6c1	1.39	Gm7658
0.37	Zcchc6	0.52	Cltb	0.58	Kdr	0.99	Tppp3	1.91	Rasl11a
0.27	Eif5	0.30	Gimap6	0.56	Plpp3	0.80	Ptprb	1.64	Cd59a
0.30	Flt1	0.42	Tspo	0.68	Dlc1	0.88	Gja4	1.36	Col4a4
0.28	Serinc3	0.44	Prr13	0.46	Ier2	0.71	Ppp1r15a	1.40	Pak1
0.37	Arpc3	0.31	Crip2	0.65	Id1	0.60	Fosb	1.62	Eif3m
0.34	Fkbp1a	0.28	Emp1	0.46	Nrp1	0.73	Cst3	1.68	Adm
0.29	Tax1bp1	0.31	Gpx1	0.54	Hsp25-ps1	0.88	Id1	1.38	Synpo
0.34	Sfpq	0.42	Atp5j	0.43	Dusp1	0.59	S100a4	1.62	Itm2b
0.29	Ywhab	0.32	Nupr1	0.56	Hbegf	0.80	Hexim1	1.52	Tmsb4x
0.26	Gabarap	0.29	Pi16	0.63	Adamts1	0.66	H2-K1	1.43	Tmsb10
0.35	Top1	0.28	Rpl11	0.51	Gng11	1.26	Fabp4	1.44	Palld
0.33	Map7d1	0.28	S100a11	0.51	2410006H16Rik	0.70	Rhob	1.54	Vegfa
0.38	Kitl	0.34	Cd200	0.48	Grasp	0.70	Heg1	1.34	Hist1h2bc
NA	NA	0.29	Klf3	0.43	Ece1	0.71	Sat1	1.43	Golim4
NA	NA	0.30	Rpl30	0.44	Gm6368	0.55	Aplp2	1.37	Mpp5
NA	NA	0.27	Rpl7	0.50	Nr4a1	0.82	Ehd4	1.40	Tspan13
NA	NA	0.34	Rab11a	0.48	Hexim1	0.57	Ier3	1.43	Thsd7a
NA	NA	0.28	Adgrf5	0.42	Pcna	0.60	Btg2	1.35	Robo2
NA	NA	0.27	Ly6e	0.46	Tek	0.70	Gadd45b	1.35	Lpin2
NA	NA	0.28	Gm10689	0.36	Ncl	0.72	Gadd45g	1.47	Fnbp1l



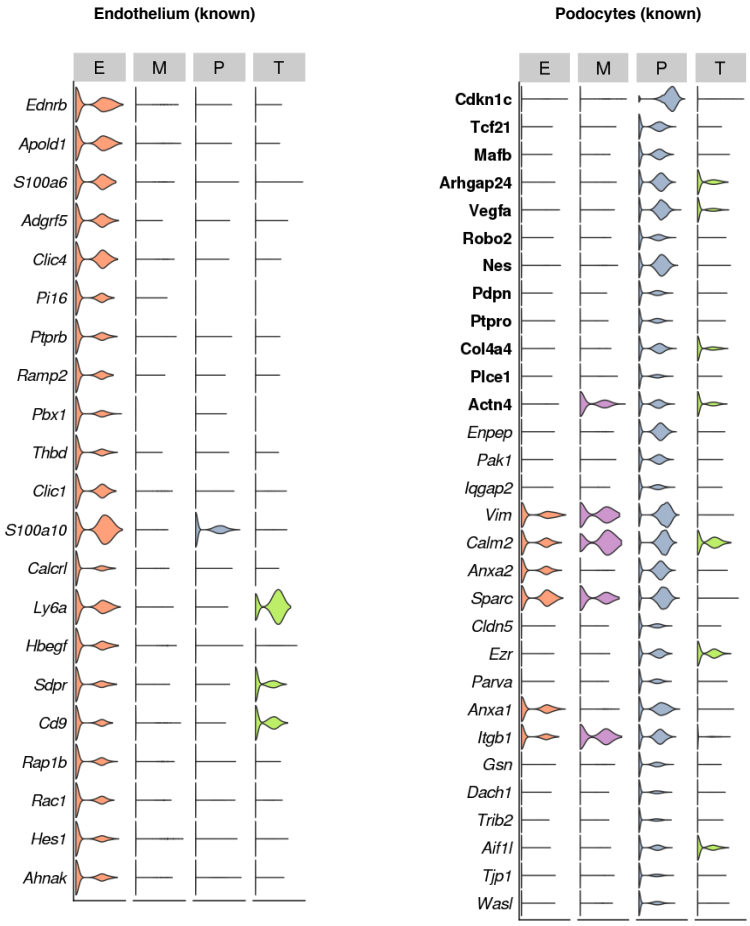
Supplemental Table 3. Top variable genes per cluster

Podocyte subclusters

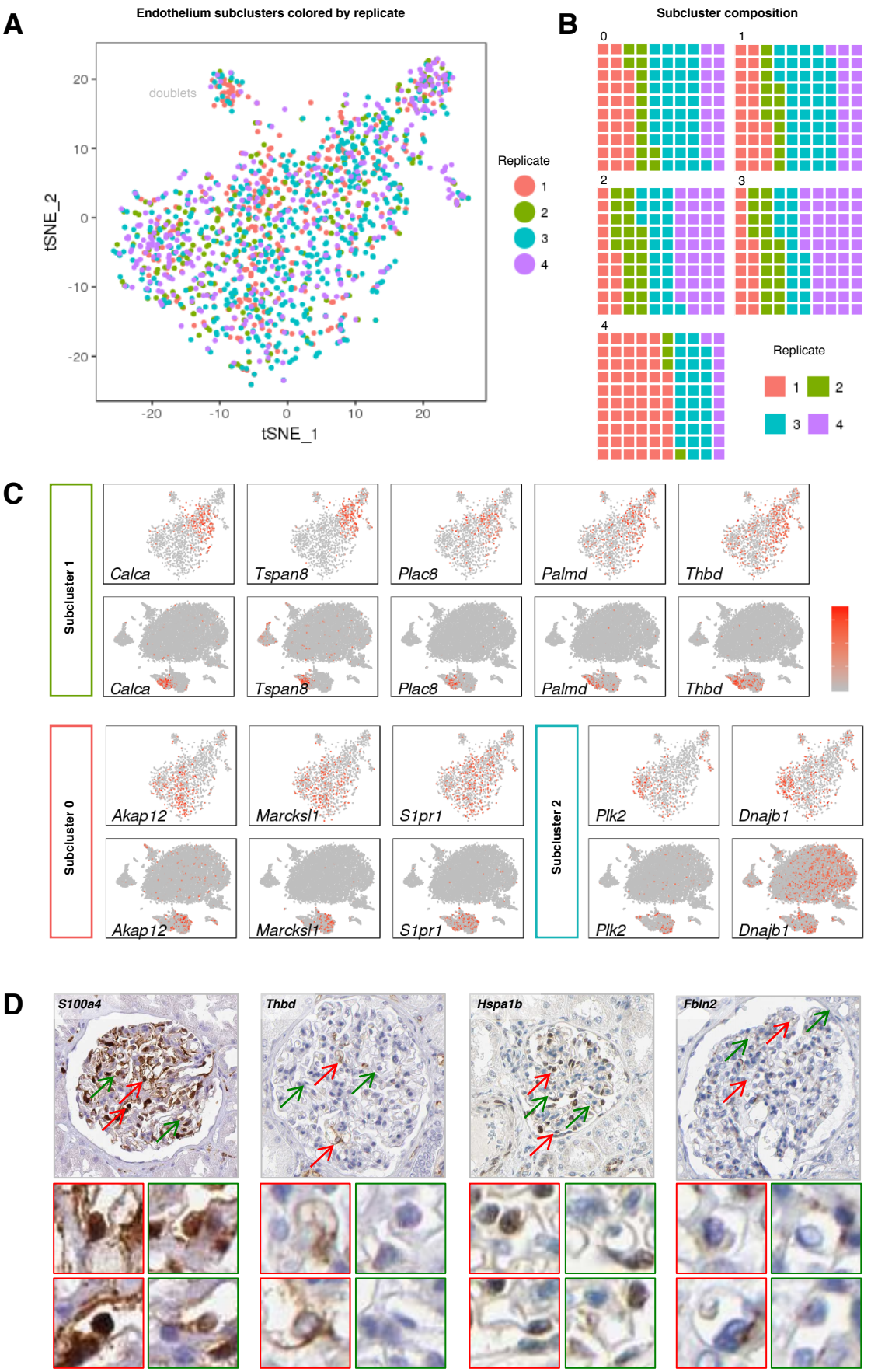
0.avg_logFC	0.gene	1.avg_logFC	1.gene	2.avg_logFC	2.gene	3.avg_logFC	3.gene	4.avg_logFC	4.gene	5.avg_logFC	5.gene
0.66	Gm37376	0.50	Tnfrsf12a	0.42	Col4a3	3.37	Gm23935	1.84	Hbegf	2.40	Gm10800
0.77	Gm26669	0.36	Nupr1	0.26	Jun	2.50	Lars2	1.83	Lsp1	1.80	Gm10801
0.61	Syn3	0.39	Ifrd1	0.28	Podxl	2.01	Gm15564	1.71	Pla2g7	0.37	Gm37376
0.40	Gm42418	0.44	Anxa1	0.33	Hspa5	1.11	Neat1	1.30	Adamts1	0.43	mt-Nd6
0.36	Ifrd1	0.30	Nfkbia	0.25	Plat	0.42	mt-Rnr2	0.95	Cald1	0.33	Zcchc6
0.36	Klf6	0.30	Gm42418	0.29	Dpp4	0.64	Podxl	0.74	Ctgf	0.26	Bclaf1
0.51	Zfp36l1	0.26	Rps27	0.26	Hsp90b1	0.70	Sema3g	0.51	Tsc22d1	0.28	Gm26669
0.30	Nfkbia	0.33	Ier3	0.27	Enpep	0.54	Enpep	0.64	Cyr61	NA	NA
0.39	Lmna	0.29	Rpl37	0.25	Sema3g	0.59	Nphs1	0.50	B2m	NA	NA
0.47	Gas1	0.35	Rasl11a	0.30	Thsd7a	0.49	mt-Rnr1	0.68	Slc6a6	NA	NA
0.55	Foxp1	0.37	Gm26532	0.25	Calr	0.55	Npnt	0.33	Atf3	NA	NA
0.51	Rn18s-rs5	0.32	Gadd45b	0.31	Pdia4	0.51	Dpp4	0.26	Vim	NA	NA
0.63	Kcnq1ot1	0.29	Rpl22l1	0.25	Htra1	0.51	mt-Nd5	0.36	Cystm1	NA	NA
0.49	Arl13b	0.25	Rpl41	0.27	Aplp2	0.45	Plat	0.47	Cryab	NA	NA
0.27	Btg2	0.28	Hspe1	0.28	Sema3e	0.40	Vegfa	0.39	Tagln2	NA	NA
0.41	Arid5b	0.26	Adm	0.27	Ctnna1	0.46	Col4a3	0.41	Chpt1	NA	NA
0.31	Btg1	0.26	Dnaja1	0.26	Myo1e	0.54	Col4a4	0.29	Gadd45b	NA	NA
0.39	Nktr	0.31	Cdk2ap2	0.25	Scd2	0.53	Thsd7a	0.34	Sec62	NA	NA
0.35	Myadm	0.28	Hspb1	NA	NA	0.36	Sparc	0.35	Rpl18a	NA	NA
0.35	BC005537	0.35	Crip1	NA	NA	0.57	Ptpro	0.26	H3f3b	NA	NA
0.33	Arl4a	0.28	Maff	NA	NA	0.35	mt-Cytb	0.37	H2-K1	NA	NA
0.32	Rb1cc1	0.26	Ddit3	NA	NA	0.53	Dag1	0.29	Gabarapl1	NA	NA
0.26	Luc7l3	NA	NA	NA	NA	0.39	mt-Nd1	0.36	Mtdh	NA	NA
0.33	Rock1	NA	NA	NA	NA	0.43	mt-Nd2	0.25	Ppp1r15a	NA	NA
0.29	Erdr1	NA	NA	NA	NA	0.37	mt-Nd4	0.30	Rpl38	NA	NA
0.27	Zbtb20	NA	NA	NA	NA	0.52	Alcam	0.40	Tuba1a	NA	NA
0.27	Rock2	NA	NA	NA	NA	0.37	Itgb1	0.26	H2-D1	NA	NA
NA	NA	NA	NA	NA	NA	0.48	Sema3e	0.30	Arid5b	NA	NA
NA	NA	NA	NA	NA	NA	0.44	Npr3	0.30	Gm26669	NA	NA
NA	NA	NA	NA	NA	NA	0.49	Efnb1	0.29	Csrp2	NA	NA
NA	NA	NA	NA	NA	NA	0.52	Angptl2	0.28	Loxl2	NA	NA
NA	NA	NA	NA	NA	NA	0.44	Itgb5	0.27	Litaf	NA	NA
NA	NA	NA	NA	NA	NA	0.47	Pros1	0.28	Tspan3	NA	NA
NA	NA	NA	NA	NA	NA	0.47	Ddn	0.34	Hnrnph1	NA	NA
NA	NA	NA	NA	NA	NA	0.45	C1qtnf1	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.42	Plod2	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.40	Pdpn	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.40	P3h2	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.39	Cldn5	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.48	Aebp1	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.47	Scd2	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.40	Magi2	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.43	Bcam	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.36	Aplp1	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.35	Pdia6	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.47	Ptprd	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.36	Dst	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.35	F2r	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.41	Tmbim6	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	0.45	Ifngr1	NA	NA	NA	NA



Supplemental Figure 2

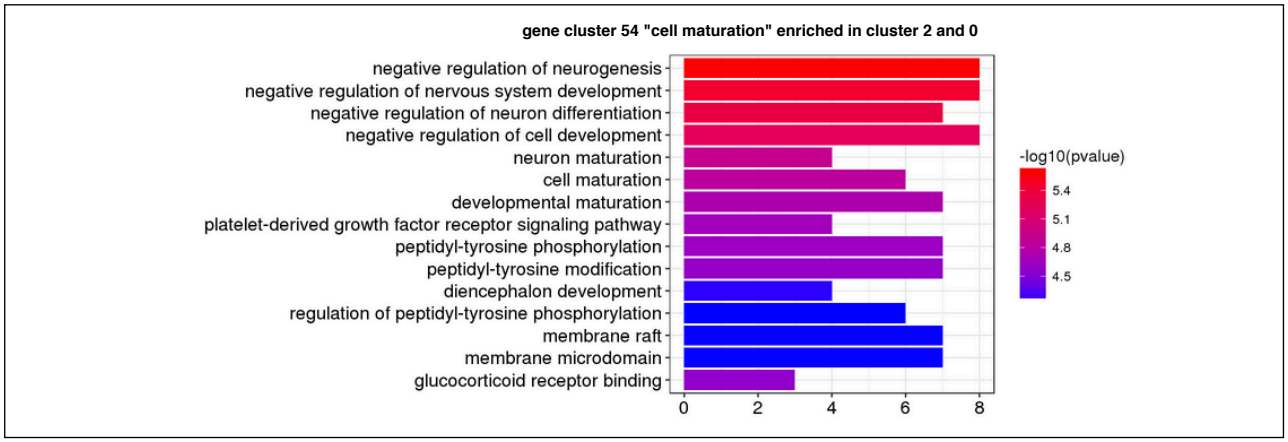


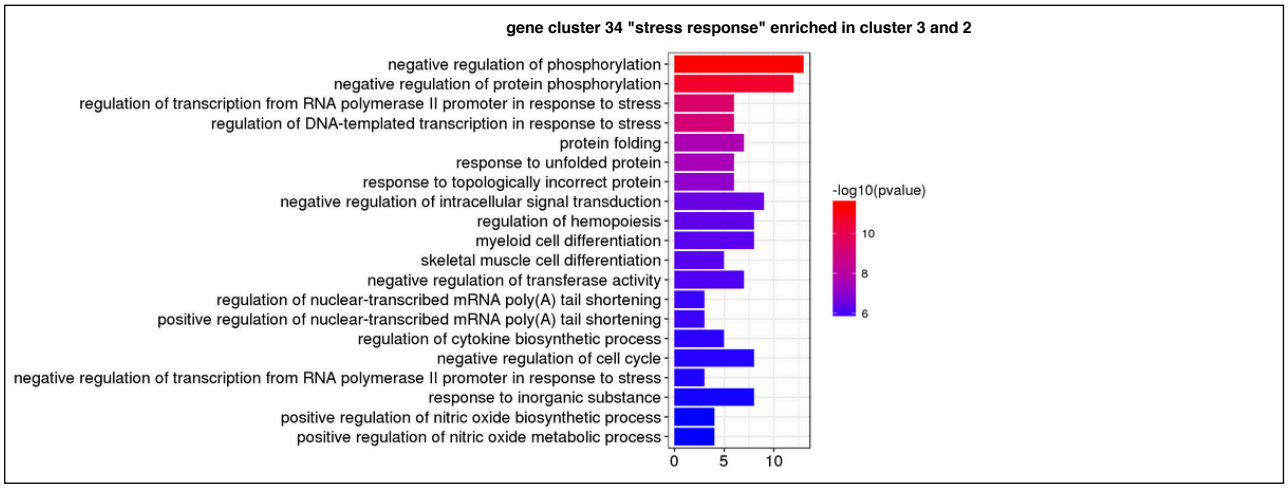
Supplemental Figure 3

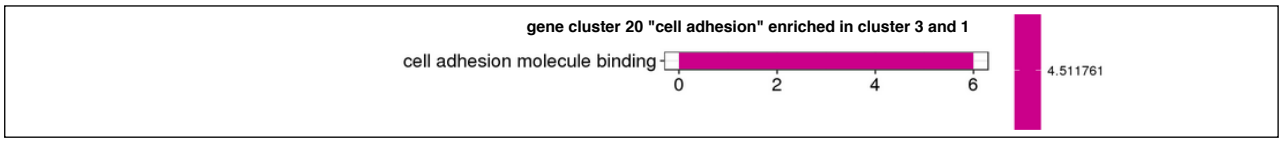


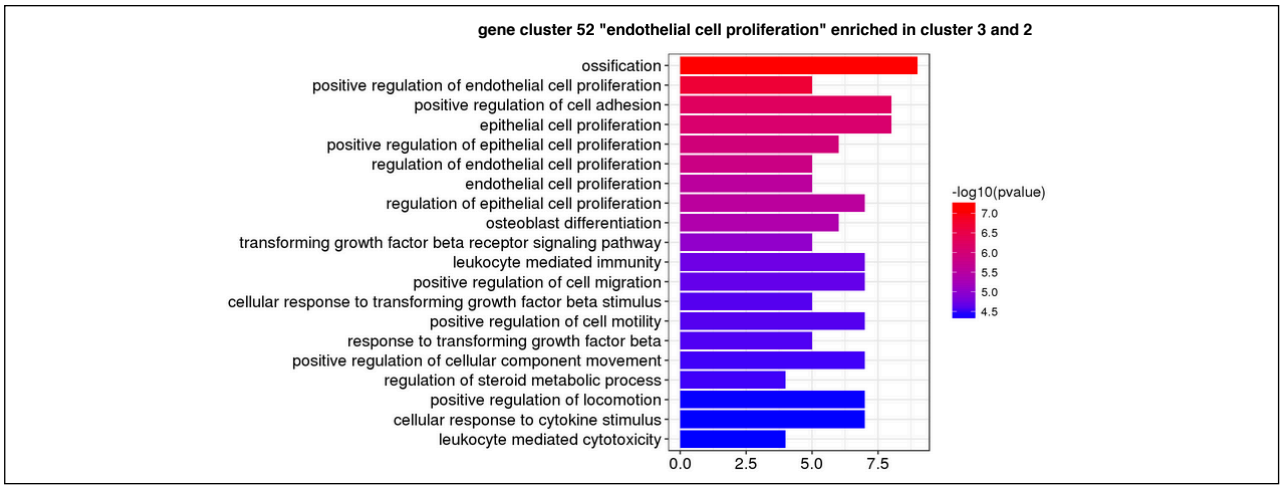
Supplemental Figure 3 continued

E





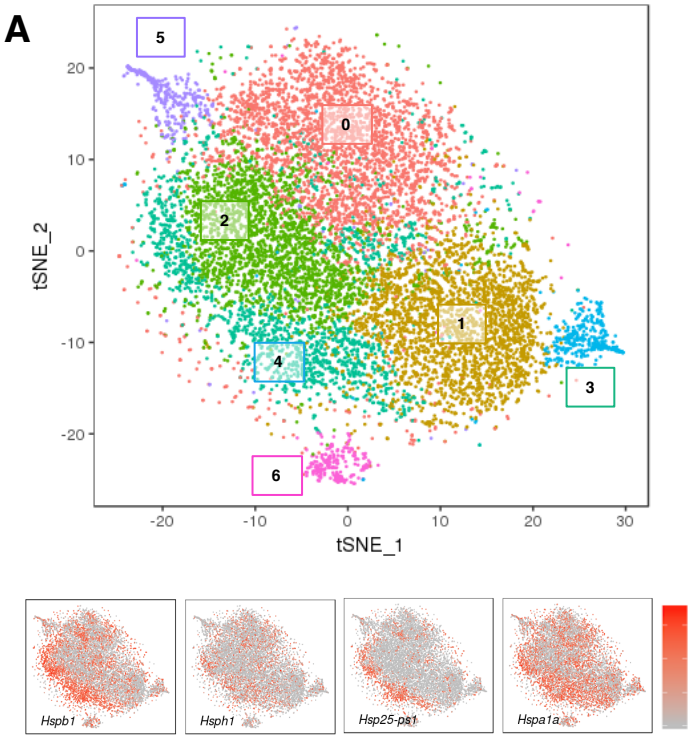


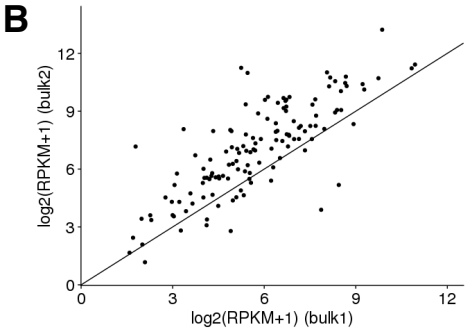


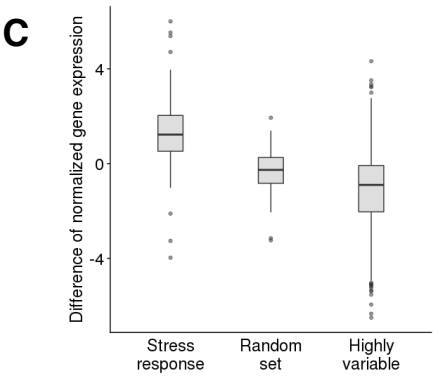
Supplemental Table 4. (Related to Fig. 3D and Suppl. Fig. 3E) Gene sets underlying the gene ontology (GO) term analysis for glomerular endothelium subclusters Barplots show the top 20 term which passed a p-value cutoff of 0.01

	geneCluster.54 enri	geneCluster.34	geneCluster.20 e	geneCluster.52 enriched in subcluster 3
1	Akap12	2410006H16Rik	1700025G04Rik	Aggf1
2	App	Atf3	Adam15	Arhgap44
3	Arap2	Btg2	Atox1	Arl15
4	Arf4	Cebpd	Calca	Art3
5	Arrdc3	Dnaja1	Calcrl	AW112010
6	Atp1a1	Dnajb1	Cd93	Bmp2
7	BC005537	Dusp1	Crip1	Casp1
8	Cd14	Egr1	Cyb5r3	Cbr2
9	Cd24a	Fos	Edn1	Cdh13
10	Cd300lg	Fosb	Ehd4	Cgnl1
11	Cdk11b	Gadd45b	Emp3	Cldn5
12	Cdkn1a	Gas5	Fam107a	Cox20
13	Crem	Gata2	Fkbp3	Ctsh
14	Csrnp1	Gm12346	Fxyd6	Cxcl1
15	Ctla2a	Gm15542	Gja4	Cxcl2
16	Ctsl	Gm6368	Ifi27l2a	Dhh
17	Cyp4b1	Hexim1	Itgb1	Dst
18	Ddx3y	Hsp25-ps1	Ly6a	Egfl8
19	Dnajb9	Hsp90aa1	Ly6e	Fbln2
20	Ednrb	Hsp90ab1	Mprip	Fbln5
21	Egfl7	Hspa1a	Palmd	Flrt1
22	Ehd3	Hspa1b	Plac8	Ganc
23	Eif1	Hspa5	Plec	Gbp3
24	Eif1-ps1	Hspa8	Polr2e	Gm4759
25	Eif4a-ps4	Hspb1	Ptrf	H2-Ab1
26	Eif4a1	Hsph1	S100a4	Htra1
27	Emcn	Icam1	Serpinb1a	Igf2
28	Ets2	Id1	Slc6a6	Igfbp4
29	Fam167b	Id2	Ssu2	Jag1
30	Gm10157	Id3	Thbd	Kcne3
31	Gm20568	Ier2	Tm4sf1	Ltbp4
32	Gmfb	Ier3	Tmsb4x	Mgp
33	Gngt2	Irf1	Tppp3	Mgst1
34	Hbegf	Jun	Tspan8	Npnt
35	Hes1	Junb	Tspo	Nuak1
36	Ifrd1	Klf2	Vim	Pir
37	Igfbp5	Klf4		Prss23
38	Ikbkb	Nfkbia		Pthlh
39	Ipmk	Nr4a1		Rsad2
40	Irf8	Phlda1		Stat5a
41	Isy1	Plk2		Stc1
42	Kdr	Ppp1r15a		Stmn2
43	Litaf	Rhob		Top2b
44	Lmcd1	Sertad1		Vcam1
45	Marcksl1	Sgk1		Vegfc
46	Mat2a	Socs3		Wfdc1
47	Midn	Tob1		Wwp1
48	mt-Cytb	Tsc22d1		
49	mt-Nd1	Ubc		
50	Mt1	Zfas1		
51	Myof	Zfp36		
52	Nhp2l1			
53	Nktr			
54	Nostrin			
55	Nr4a2			
56	Nrp1			
57	Ntn4			
58	Paip2			
59	Pbx1			
60	Plat			
61	Plaur			
62	Plpp1			
63	Plpp3			
64	Prkab1			
65	Ramp3			
66	Rell1			
67	Sec62			
68	Slc43a3			
69	Slfn5			
70	Smarca5			
71	Spry4			
72	Stat3			
73	Timeless			
74	Tiparp			
75	Tpm3			
76	Tpm3-rs7			
77	Tra2a			
78	Uqcrq			

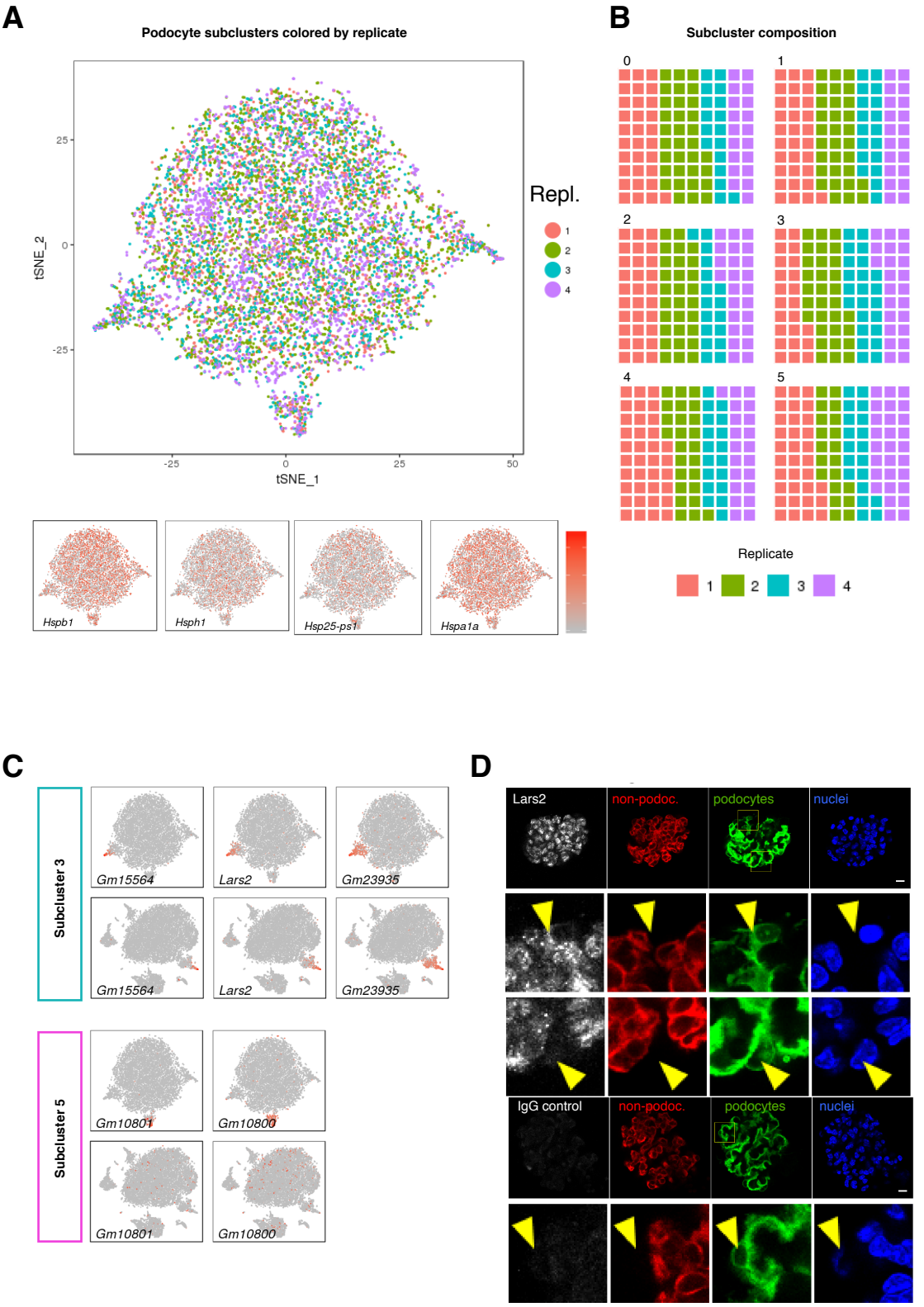
Supplemental Figure 4





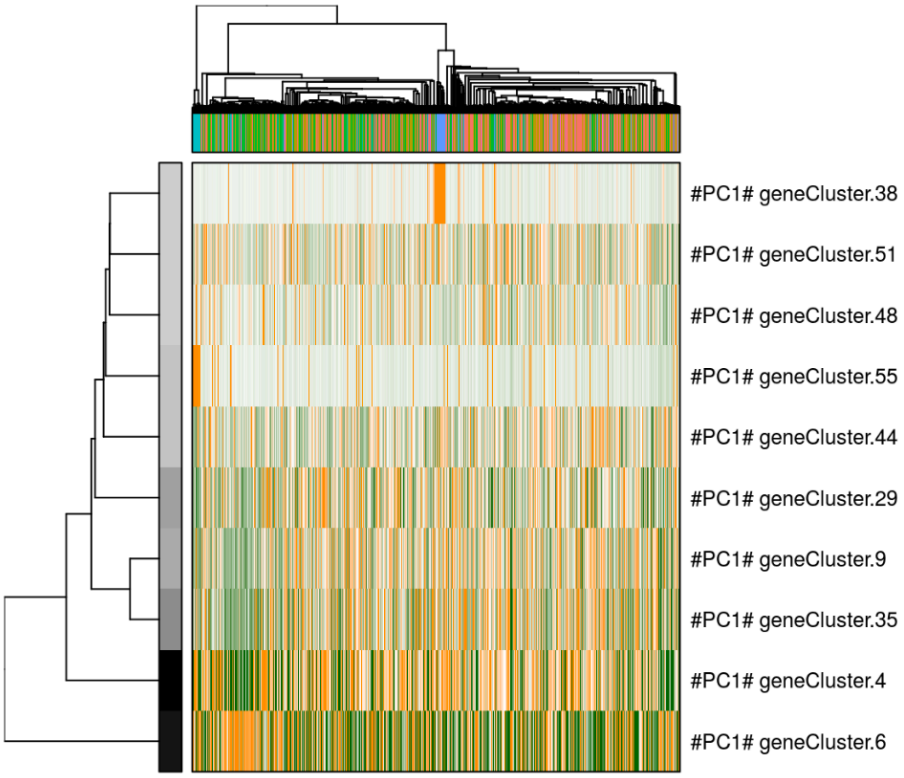


Supplemental Figure 5



Supplemental Figure 6

A



B

gene cluster 38

Aldh1a1	Pdlim4
Bmp4	Pla2g7
Cd200	Plppr4
Csflr	Prss23
Dkk3	Ptgis
Efemp1	Rasl11b
Fam180a	Rbm1
Hey2	Smim5
Igfbp2	Tgm2
Lsp1	Tmem176b
Mustn1	Wwc1
Ntn4	

gene cluster 55

Gm15564
Gm23935
Lars2

Supplemental Figure and Table legends *(order as they appear in the main text)*

Supplemental Figure 1. (Related to Fig. 1) **Quality of glomeruli single-cell data and comparisons with bulk mRNAseq data**
(A) Representative brightfield microscopic images of isolated mouse glomeruli showing high purity of glomeruli. Scale bar corresponds to 100 µm. Magnified areas indicate a tubular fragment (left), a decapsulated glomerulus with protruding vasculature (middle) and an avascular, decapsulated glomerulus (right). Quantification of 123 structures in a 10µl droplet of a glomerular preparation: tubulular fragments 13% (n=16), vascular glomeruli 13% (n=16), avascular glomeruli 74% (n=91), Bowman's capsules 0% (n=0)
(B) Distribution and median of the number of genes and UMIs detected per cell in four separate Drop-seq runs corresponding to independent biological replicates (left panel) or to a pool of replicates (right panel). Cells were filtered to express >250 genes and >350 UMIs. In the pool, cell doublets were removed computationally (see Methods and Suppl. Table 1 for median number of genes, UMIs and reads per cell).
(C) tSNE plot from Fig. 1B with each cell now coloured according to biological replicate. The batch effect visible in the podocyte cluster was regressed out for subclustering analysis (Fig. 4; Suppl. Fig. 5A, B).
(D) Heat map indicating that samples cluster by cell type, not by experimental replicate. The fraction of each cell type per replicate is indicated in the bar plot on top.
(E) Pairwise correlations between Drop-seq single-cell data and bulk mRNA seq samples of similar depth generated in this study. "bulk1" corresponds to glomerular lysate (before single-cell dissociation) and "bulk2" corresponds to single-cell suspensions prepared by enzymatic digestion followed by live cell sorting by flow cytometry. The analysis showed high correlations in general, but also indicated that single-cell data correlate slightly better with the bulk data from cells that underwent single-cell dissociation (circles).
(F) Scatterplots showing the expression in the "bulk1" and "bulk 2" samples for the 50 most variable genes per cell type identified in the single-cell RNAseq data. The data suggest that podocytes were overrepresented in the "bulk2" sample, indicating that cell dissociation introduced a bias. Endothelial and mesangial cells showed a higher dispersion with the opposite tendency, while tubules were not affected.
(G) Boxplots of expression ratios in the bulk data for the 50 most variable genes per cell type (corresponding to (F)). A distribution along the horizontal line would indicate absence of bias.
(H) Pairwise correlations between cell types identified by single-cell data generated in this study (pre-fix "ds") and published bulk mRNA-seq data sets. The analysis showed the highest correlations with the corresponding cell type, supporting cell type assignments in the single-cell data. Scale for the heatmap is indicated. R = Pearson correlation efficient.
mRNAseq data were from Brunskill et al. 2014[12]: gsm1585030 (proximal tubules), gsm1585043 (MafB podocytes), gsm1585044 (mesangial cells), gsm1585047 (endothelial cells); from Kann et al.[1]: gsm156360 (podocytes).

Supplemental Table 1. (Related to Suppl. Fig. 1B). **Median number of genes, transcripts (UMIs) and reads per cell**
This table lists the number of uniquely mapped reads per cell, the number of cells, the number of detected genes and the number of transcripts per cell (defined as unique molecular identifiers (UMIs). Numbers are given for data before or after cell doublet removal.

Supplemental Table 2. (Related to Suppl. Fig. 1B). **Marker genes for each of the five major cell types to identify putative cell doublets**
Cell-type specific genes as listed in this table were used to identify putative cell doublets (see Online Methods for details).

Supplemental Table 3. (Related to Suppl. Fig. 1B, Fig. 3A and Fig. 4A). **Top variable genes per cell cluster or subcluster**
This table consists of three subtables which list the 50 most highly variable genes in each cell cluster as determined by the function 'FindAllMarkers' in Seurat[6]. 'Full dataset' corresponds to Fig. 1B (cell type clusters). 'Endothelium' corresponds to Fig. 3A (endothelium subclusters). 'Podocytes' corresponds to Fig. 4A (podocyte subclusters).

Supplemental Figure 2. (Related to Fig. 2) **Additional markers of glomerular cells**
Distribution and relative expression of individual highly variable genes (violin plots) in endothelium and podocytes. Established markers (bold) and markers described in the literature (*italics*).

Supplemental Figure 3. (Related to Fig. 3) **Subclustering of glomerular endothelium**
(A) 2D representation corresponding to the plot in Fig. 3A. Replicate Drop-seq experiments 1 to 4 contributed to all clusters.
(B) Waffle plots show that each cluster contained cells from all four replicates, while the composition of cluster 4, identified as residual cell doublets, was biased. A square represents 1% of the total number of cells in a cluster.
(C) Expression of markers colored based on normalized expression levels. Upper panels correspond to subcluster tSNE plot shown in (A), lower panels to the tSNE plot of the whole data set as shown in Fig. 1B.
(D) Immunohistochemistry images obtained from Human Protein Atlas (HPA)[10] support that endothelial subcluster marker proteins *S100a4*, *Thbd*, *Hspa1b*, and *Fbln2* may be expressed heterogeneously in human glomerular endothelium. Area shown (500 x 500 pixels = 200 μm^2) corresponds to glomeruli taken from a larger HPA image. Red arrows point to endothelial cells as identified by intracapillary nuclei with strong (*S100a4*) or positive (*Thbd*, *Hspa1*, *Fbln2*) marker expression. Green arrows point to endothelial cells with weak (*S100a4*) or no (*Thbd*, *Hspa1*, *Fbln2*) marker gene expression. Magnified areas (50 x 50 pixels = 20 μm^2) are shown below with red squares representing strong/positive expression and green squares weak or no expression.
(E) Gene ontology (GO) terms corresponding to heat map shown in Fig. 3C. The heatmap indicates that 4 endothelial subclusters (0: red; 1: green; 2: blue; 3: violet) show distinct, overrepresented gene activation patterns (see also Suppl. Fig. S3C): Cluster 1 and 3 (gene set 20 terms related to “cell adhesion”), Cluster 2 and 0 (gene set 54 terms related to “cell maturation”), cluster 2 and 3 (gene set 34 terms related to ”stress response”), and cluster 3 (gene set 52 terms related to “endothelial cell proliferation”; see panel E). p values were calculated by hypergeometric testing (Fisher's exact test). Corresponding gene clusters are listed in Suppl. Table 4.

Supplemental Table 4. (Related to Suppl. Fig. 3D and Suppl. Fig. 3E). **Gene sets underlying the gene ontology (GO) term analysis for glomerular endothelium subclusters**
This table lists the sets of genes, expressed in endothelium subclusters, that underlie the GO analysis shown in Fig. 3D and Suppl. Fig. 3E.

Supplemental Figure 4. (Related to Fig. 4) **Stress response in podocytes**
(A) Upper panel: 2D tSNE representation of podocyte subclustering before correcting for tissue-dissociation-induced stress response genes[7]. Subcluster 4 showed upregulation of stress response genes. Lower panel: Expression of stress response markers colored based on normalized expression levels (grey: low, red: high).
(B) Stress response genes, known to become transcriptionally induced by tissue-dissociation[7], show higher expression in bulk mRNA-seq samples from glomerular cells that

underwent prolonged enzymatic tissue dissociation ("bulk2") as compared to non-dissociated glomeruli ("bulk1"). The set of 140 genes was taken from van den Brink et al. 2017[7], Suppl. Table S5:

Actg1, Ankrd1, Arid5a, Atf3, Atf4, Bag3, Bhlhe40, Brd2, Btg1, Btg2, Ccnl1, Ccrn4l, Cebpb, Cebpd, Cebpg, Csrnp1, Cyr61, Cxcl1, Dcn, Ddx3x, Ddx5, Des, Dnaja1, Dnajb1, Dnajb4, Dusp1, Dusp8, Erf, Egr1, Egr2, Eif1, Eif5, Errfi1, Fam132b, Fos, Fosb, Fosl2, Gadd45a, Gadd45g, Gcc1, Gem, H3f3b, Hipk3, Hspa1a, Hspa1b, Hspa5, Hspa8, Hspb1, Hspe1, Hsph1, Hsp90aa1, Hsp90ab1, Idi1, Id3, Ier2, Ier3, Ier5, Il6, Irf1, Ifrd1, Itpkc, Jun, Junb, Jund, Kcne4, Klf2, Klf4, Klf6, Litaf, Lmna, Maff, Mafk, Mcl1, Midn, Mir22hg, Mt1, Mt2, Myadm, Myd88, Myc, Nckap5l, Ncoa7, Nfkbia, Nop58, Nppc, Nr4a1, Odc1, Osgin1, Oxnad1, Pcf11, Per1, Phlda1, Pnrc1, Pnp, Ppp1cc, Ppp1r15a, Pxdc1, Rassf1, Rhob, Rhoh, Ripk1, Sat1, Sbno2, Sdc4, Skil, Slc10a6, Slc38a2, Slc41a1, Socs3, Sqstm1, Srf, Srsf7, Stat3, Tagln2, Tiparp, Tnfaip3, Tnfaip6, Tpm3, Tra2a, Tra2b, Trib1, Tubb4b, Tubb6, Usp2, Irf8, Klf9, Nfkbiz, Pde4b, Rap1b, Serpine1, Srsf5, Tppp3, Ubc, Wac, Zc3h12a, Zfand5, Zfp36, Zfp36l1, Zfp36l2, Zyx.

(C) Quantitation of the tissue-dissociation-induced stress response between "bulk2" and "bulk1" mRNA-seq samples for genes with a normalized expression greater than 1 (132 out of 140). Stress-response genes showed significantly elevated expression in single, dissociated cells compared to comparable sets of random genes or highly variable genes (for details see Methods; t-test $p<2.2e-16$ for each comparison).

Supplemental Figure 5. (Related to Fig. 4) **Subclustering of podocytes**

(A) Upper panel: 2D representation of podocyte subclustering after correcting for stress response by regressing out stress response genes. Cells are colored by replicates (1 to 4). Replicates contribute to all clusters. Lower panel: Expression of stress response markers colored based on normalized expression levels (grey: low, red: high).

(B) Waffle plots showing that each cluster contained cells from all four replicates. A square represents 1% of the total number of cells in a cluster.

(C) Expression of markers colored based on normalized expression levels. Upper panels correspond to subcluster tSNE plot as shown in (Fig. 3A), lower panels to the tSNE plot of the whole data set as shown in Fig. 1B.

(D) Laser scanning confocal microscopy of glomeruli purified from kidneys of transgenic Nphs2-Cre x mT/mG double-fluorescent Cre reporter mice[11]. Podocytes are genetically marked by mG (Cre-dependent membrane-targeted GFP; green) fluorescence, while non-podocyte cell types remain mT (membrane-targeted tandem dimer Tomato; red)-positive. Top row: purified, whole glomeruli. Lars2 indicates Lars2 antibody staining (left) or IgG control (right). Second row: Yellow arrowheads points to a Lars2-positive podocyte. Third row: Yellow arrowhead points to a Lars2-negative podocyte. Scale bars correspond to 10µm. Magnified areas are 22 x 22 μm^2. IgG isotype control images showed no signal.

Supplemental Figure 6. (Related to Fig. 4) **Pathway and gene set overdispersion analysis for podocyte subclusters**

(A) Pathway and gene set overdispersion analysis (PAGODA)[8] of the podocyte subclusters shown in Fig. 4A,B. The heatmap shows distinct, overrepresented gene activation patterns only for subclusters 4 (blue) and 3 (turquoise).

(B) Gene ontology (GO) terms corresponding to the heat map in (A). Podocyte subclusters 4 and 3 express gene sets that had no associated GO terms.

References for Supplemental Material

1. Kann M, Ettou S, Jung YL, Lenz MO, Taglienti ME, Park PJ, Schermer B, Benzing T, Kreidberg JA: Genome-Wide Analysis of Wilms' Tumor 1-Controlled Gene Expression in Podocytes Reveals Key Regulatory Mechanisms. *J. Am. Soc. Nephrol. JASN* 26: 2097–2104, 2015

2. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA: Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161: 1202–1214, 2015

3. Alles J, Karaiskos N, Praktiknjo SD, Grosswendt S, Wahle P, Ruffault P-L, Ayoub S, Schreyer L, Boltengagen A, Birchmeier C, Zinzen R, Kocks C, Rajewsky N: Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* 15: 44, 2017

4. Karaiskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, Kocks C, Rajewsky N, Zinzen RP: The Drosophila embryo at single-cell transcriptome resolution. *Science* 358: 194–199, 2017

5. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* 29: 15–21, 2013

6. Satija R, Farrell JA, Gennert D, Schier AF, Regev A: Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33: 495–502, 2015

7. van den Brink SC, Sage F, Vértesy Á, Spanjaard B, Peterson-Maduro J, Baron CS, Robin C, van Oudenaarden A: Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* 14: 935–936, 2017

8. Fan J, Salathia N, Liu R, Kaeser GE, Yung YC, Herman JL, Kaper F, Fan J-B, Zhang K, Chun J, Kharchenko PV: Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods* 13: 241–244, 2016

9. Yu G, Wang L-G, Han Y, He Q-Y: clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* 16: 284–287, 2012

10. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA-K, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist P-H, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F: Proteomics. Tissue-based map of the human proteome. *Science* 347: 1260419, 2015

11. Muzumdar MD, Tasic B, Miyamichi K, Li L, Luo L: A global double-fluorescent Cre reporter mouse. *Genes. N. Y. N 2000* 45: 593–605, 2007

12. Brunskill EW, Park J-S, Chung E, Chen F, Magella B, Potter SS: Single cell dissection of early kidney development: multilineage priming. *Dev. Camb. Engl.* 141: 3093–3101, 2014