Supplementary Methods

Calculation of F-measure

The F-measure can be calculated for a given trade-off between precision and recall as

$$F_{x/y} = \frac{\left(1+\left(\frac{x}{y}\right)^2\right)\times(P\times R)}{\left(\frac{x}{y}\right)^2\times P + R}$$ where x/y times as much weight is given to recall as precision. We set $x=1$ and $y=5$. This weighs precision five times as much as recall when selecting the appropriate cut-off for experimental and clinical validation.

Correlation based prediction of cell-lineage-specific genes

We used a correlation-based strategy for predicting cell-lineage specific expression as a baseline method. We used the Sleipnir library for functional genomics to calculate a z-transformed correlation of the gene expression for all genes in the micro-dissected human kidney data. We then ranked genes by their average z-transformed correlation to all genes in the podocyte gold standard. While the SVM was able to successfully separate different glomerular compartments (Fig. 2), the baseline correlation approach did not achieve useful separation (SI Fig. 1).

Gene expression data extraction from the GUDMAP

Briefly, podocytes, mesangial cells and endothelial cells were isolated from kidneys of adult Mafb-EGFP transgenic mice (Tg(Mafb-EGFP)79Gsat), Meis-EGFP transgenic mice (Tg(Meis1-EGFP)FO156Gsat), and Tie2-GFP transgenic mice (Tg(TIE2GFP)287Sato) respectively, using a combination of collagenase A digestion, sieving and trypsinization followed by Fluorescence Activated Cell Sorting (FACS). Affymetrix Mouse Gene 1.0 ST Arrays were used by GUDMAP to obtain gene expression data following the Affymetrix standard protocol.

The robustness *in silico* nano-dissection to poor or limited standards

To assess the robustness of our method when our lineage-specific positive markers are limited or of varying quality, we examine the performance of nano-dissection under two sets of conditions. In the first case, we subsample from our gold standard to create a smaller gold standard containing only 50-90% of the genes in our full standard (SI Fig 4). In the second case, we add 1 to 4 identically sized standards made up of randomly selected genes to our gold standard. This generates new standard sets 2-5 times the size of our gold standard. We then compare nano-dissection's performance to that of linear SVM using these sets of gold standards (SI Fig 5). In both cases, we observe that nano-dissection is robust to such perturbations and effectively separates podocyte and non-podocyte genes, even under such sub-optimal conditions.

Comparison of *in silico* nano-dissection to genome-wide GFR correlation

To evaluate whether nano-dissection provides information not available from a genome-wide GFR correlation analysis when the goal is to identify genes associated with diseases that show a cell lineage specific etiology, we performed a systematic analysis focused on hereditary glomerular disease. We curated genes associated with hereditary FSGS in prior literature and excluded any that were included within the positive gold standard for nano-dissection. These thirteen FSGS-associated genes (*APOL1*, *ARHGAP24*, *COL4A3*, *COL4A4*, *INF2*, *ITGB4*, *LAMB2*, *MYH9*, *MYO1E*, *PLA2R1*, *SCARB2*, *SMARCAL1*, and *TSPAN13*) do not overlap with our podocyte positive gold standards.

First, we evaluated, for all genes, the absolute value of their expressions' correlation with the square root of GFR measurements, termed the "GFR correlation score". The absolute value of this correlation is used because genes associated with a disease may either have a positive or negative relationship with renal function. Genes associated with hereditary FSGS (red, SI Fig. 8A) were observed to have a greater GFR correlation score than genes without demonstrated FSGS association (blue, SI Fig, 8A), but the relationship was not statistically significant (t-test $p = 0.17$). Thus a genome-wide assessment of GFR

correlation alone was not capable of identifying genes associated with this hereditary renal disease. In contrast to the results for the GFR correlation score, genes associated with FSGS were given significantly higher scores by podocyte nano-dissection (red, SI Fig. 8B) (t-test $p = 0.005$), as compared to nano-dissection scores of genes without known FSGS association (blue, SI Fig. 8B).

Figure legends for supplementary figures:

Supplementary Fig. 1. The distribution of tissue specific genes by percentile using classical correlation approach. This approach was not able to separate known podocyte genes from genes specific to the other glomerular cell-lineages and tubular cells.

Supplementary Fig. 2: Nano-dissection is applicable to non-renal cell lineages. Here we separate astrocyte (green) genes from microglia (blue) and oligodendrocyte (red) genes. Cell-type-specific standards for each type were obtained from HPRD, and nano-dissection was applied to the hand-curated compendium of brain samples available in the nano-dissection webserver (analysis URL: http://nano.princeton.edu/analysis/view/8c02f58c-10c0-4693-832b-f40a0e8c9cca/). The top 150 predictions (dotted line), as used in our podocyte study, provide a highly accurate separation of astrocytes from the other lineages.

Supplementary Fig. 3: Nano-dissection is applicable to non-renal cell lineages, even on whole-organism data compendia. Here we separate fibroblast (red) genes from keratinocyte (green) and melanocyte (blue) genes. Cell-type-specific standards for each type were obtained from HPRD, and nano-dissection was applied to the entire human gene-expression compendium available in the nano-dissection webserver (analysis URL: http://nano.princeton.edu/analysis/view/f8e6845c-ed51-433e-b558-5782ac885154/). The top 150 predictions (dotted line), as used in our podocyte study, provide a highly accurate separation of fibroblasts from the other lineages.

Supplemental Fig. 4: Nano-dissection's performance is robust to limited gold standards and those of varying quality. The observed performance is shown at 100% of standard (blue dot). The performance of 50 trials sub-sampled to 50-90% of the existing standard size is shown to the left of 100% and red bars indicate 95% confidence intervals.

Supplementary Fig. 5: Iterative nano-dissection approach is robust to low quality gold standards, whereas SVM is not. We evaluate both our method and the linear SVM in the context of adding a contaminating random gold standard. The plot shows the AUC of each approach when a random standard of 1, 2, 3, or 4 times the size of our gold standard is added as an additional standard. The iterative nano-dissection approach identifies that the standard is useless and avoids it, while the SVM's performance deteriorates substantially. The shaded regions show 95% confidence intervals (nonparametric bootstrap with 10,000 resamples) around the mean.

Supplementary Fig. 6: Nano-dissection is applicable to non-podocyte renal lineages. Here we separate mesangial (light blue) genes from endothelial (dark blue), podocyte (green) and tubular (red) genes. Nano-dissection was applied to the renal microdissection compendium available in the nano-dissection webserver (analysis URL: http://nano.princeton.edu/analysis/view/Mesangial-nano-dissection-hUjTVxrPBnP38T2c/).

Supplementary Fig. 7: Pseudocode for our algorithm showing the nano-dissection decision rule. The array d[] represents the results of the chosen classification algorithm, here the distance to the hyperplane from the SVM. The lineage-specific validation procedure is used when the validation standard is specific to the lineage of interest, which also occurs when no external validation standard is available. The compartment validation procedure is used when the validation standard includes lineages expected to occur in both the positive and negative standards. Also see the Methods section for an in depth description.

Supplementary Fig. 8: Genes associated with FSGS do not have a significantly different correlation to GFR (A, red bar) than genes with unknown status (A, blue bar), but FSGS associated genes do have significantly higher nano-dissection scores (B, red bar) than genes not known to be associated with FSGS (B, blue bar). Error bars indicate 95% confidence intervals as assessed by 1000 bootstrap samples.

Supplementary Table 1. Positive standards for podocytes selected by literature search.

Supplementary Table 2. Negative standards for podocytes selected by literature search.

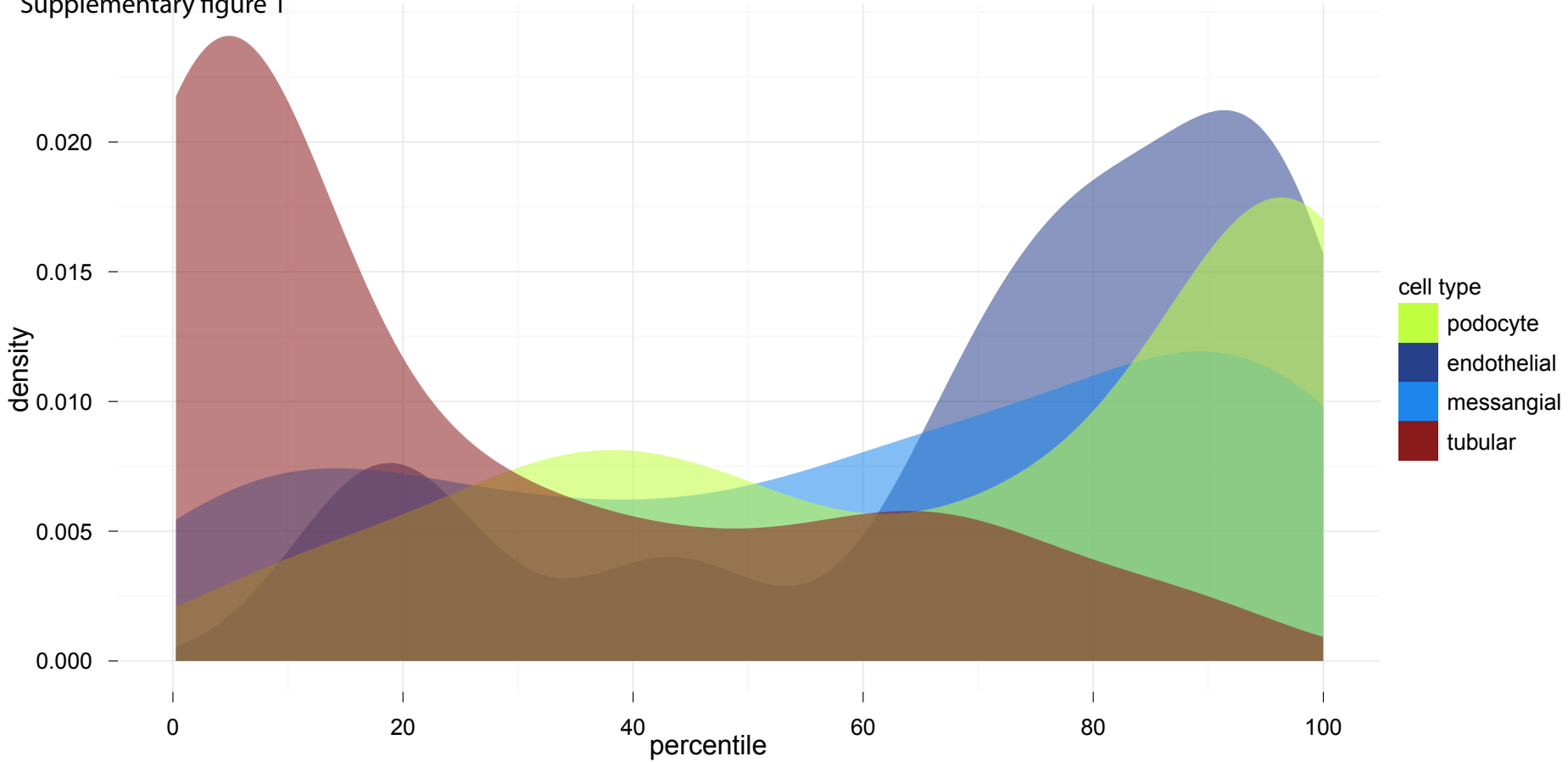
Supplementary Table 3. Top 136 predicted genes by *in silico* nano-dissection approach.

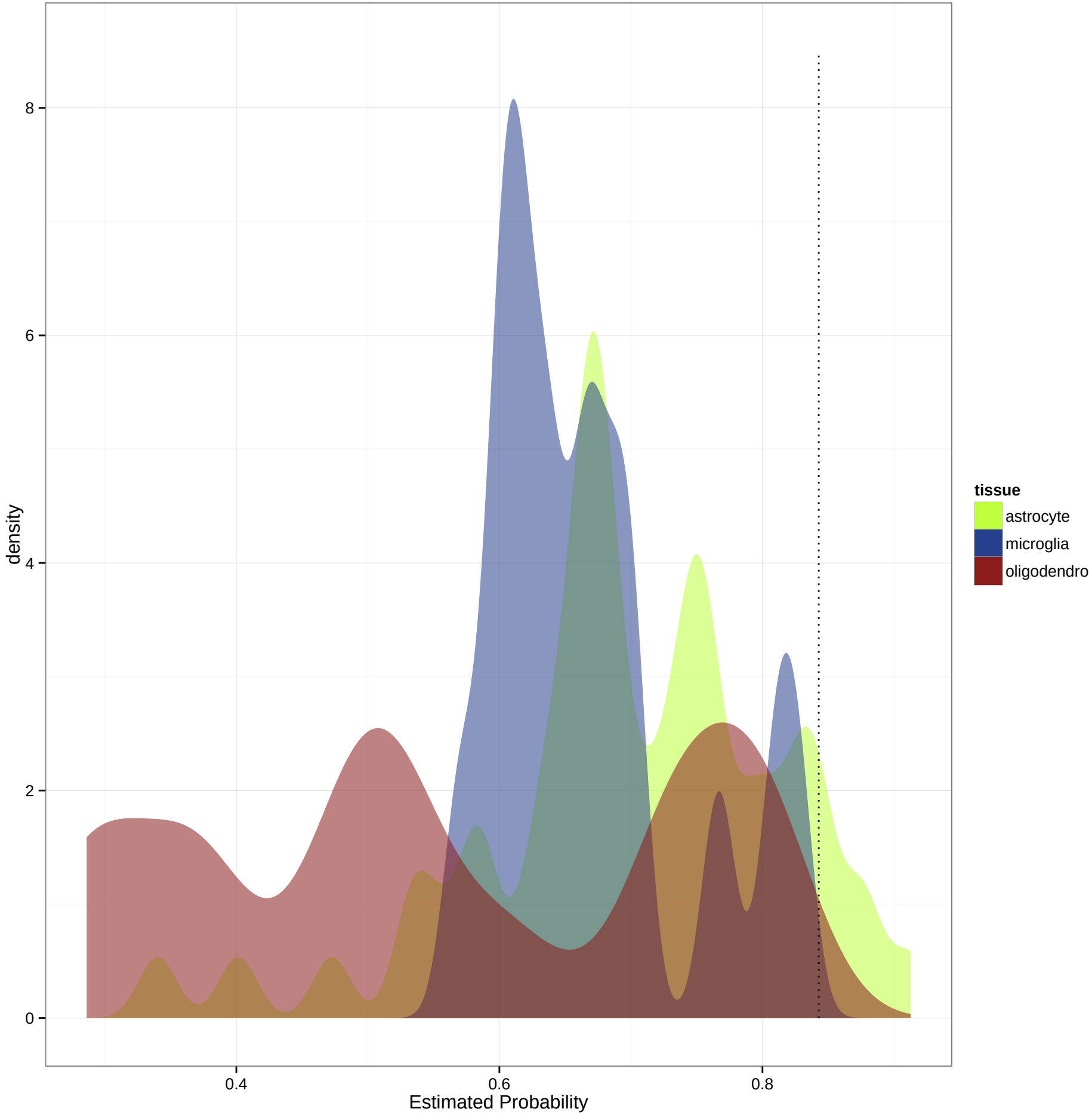
Supplementary Table 4. GUDMAP gene expression datasets have been used for this study.

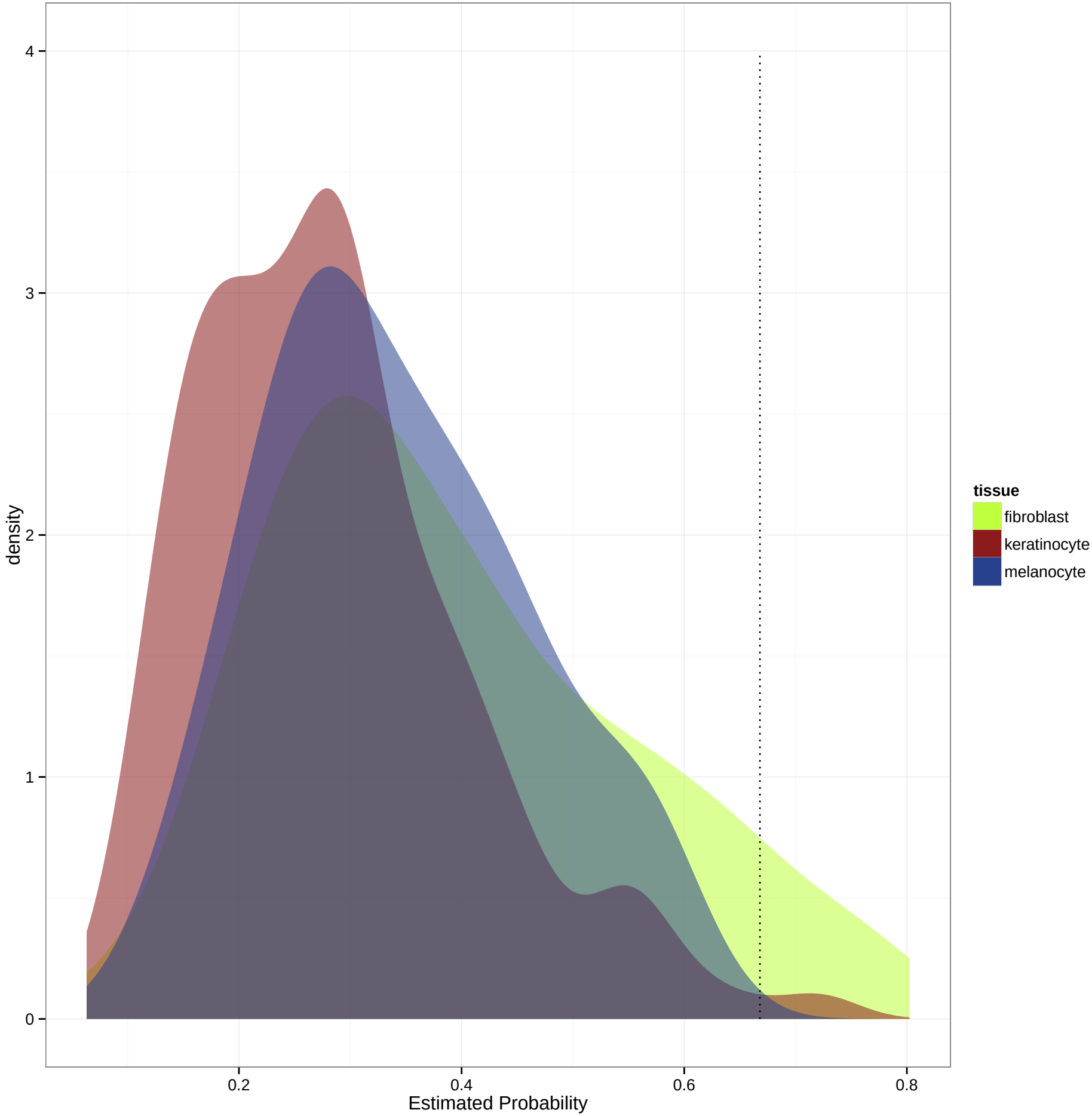
Supplementary Table 5. 102 genes exhibit significantly higher expression in podocytes in comparison with the other two major cell types in mouse glomerulus.

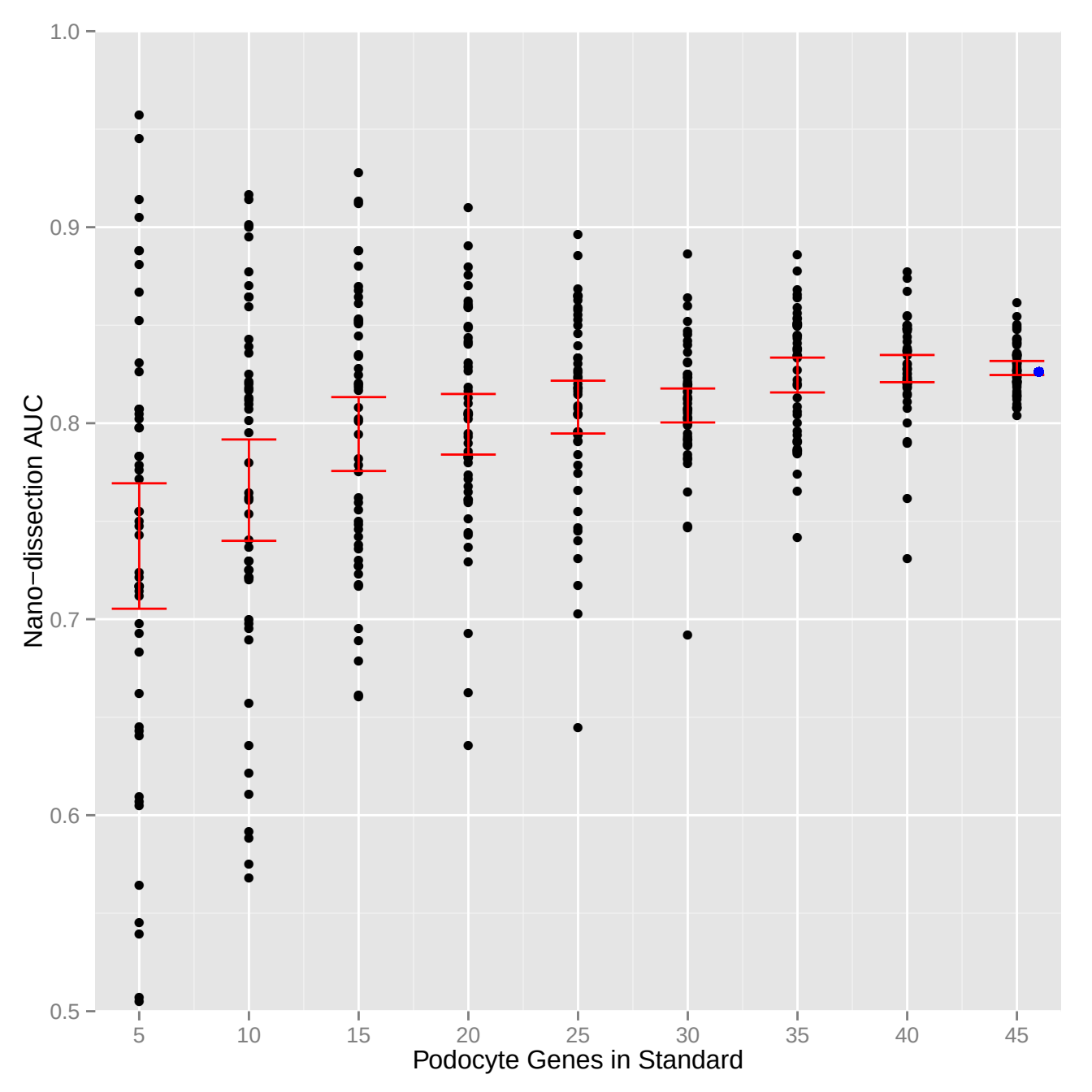
Supplementary Table 6. Demographics of 139 CKD patients.

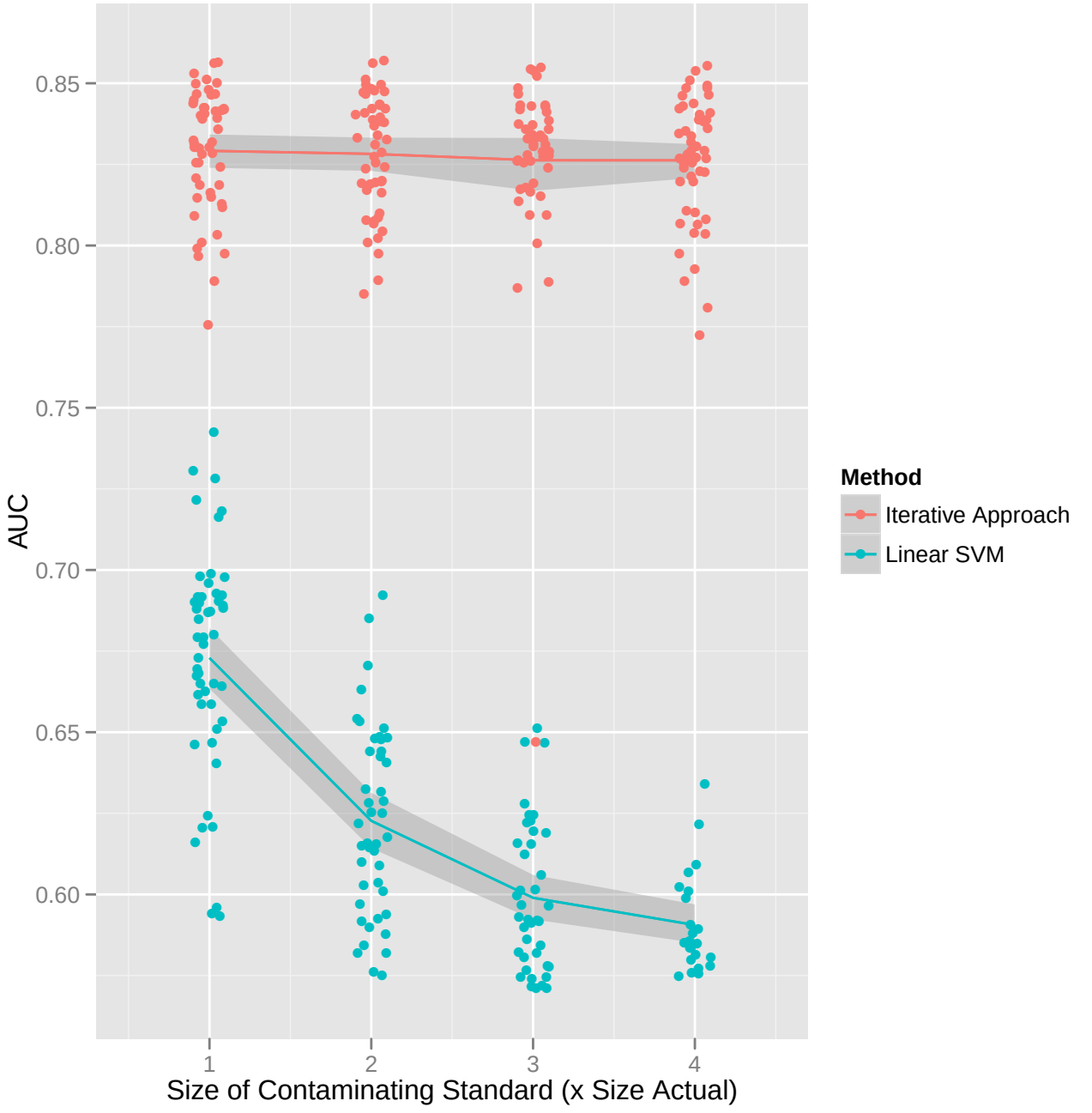
Supplementary figure 1

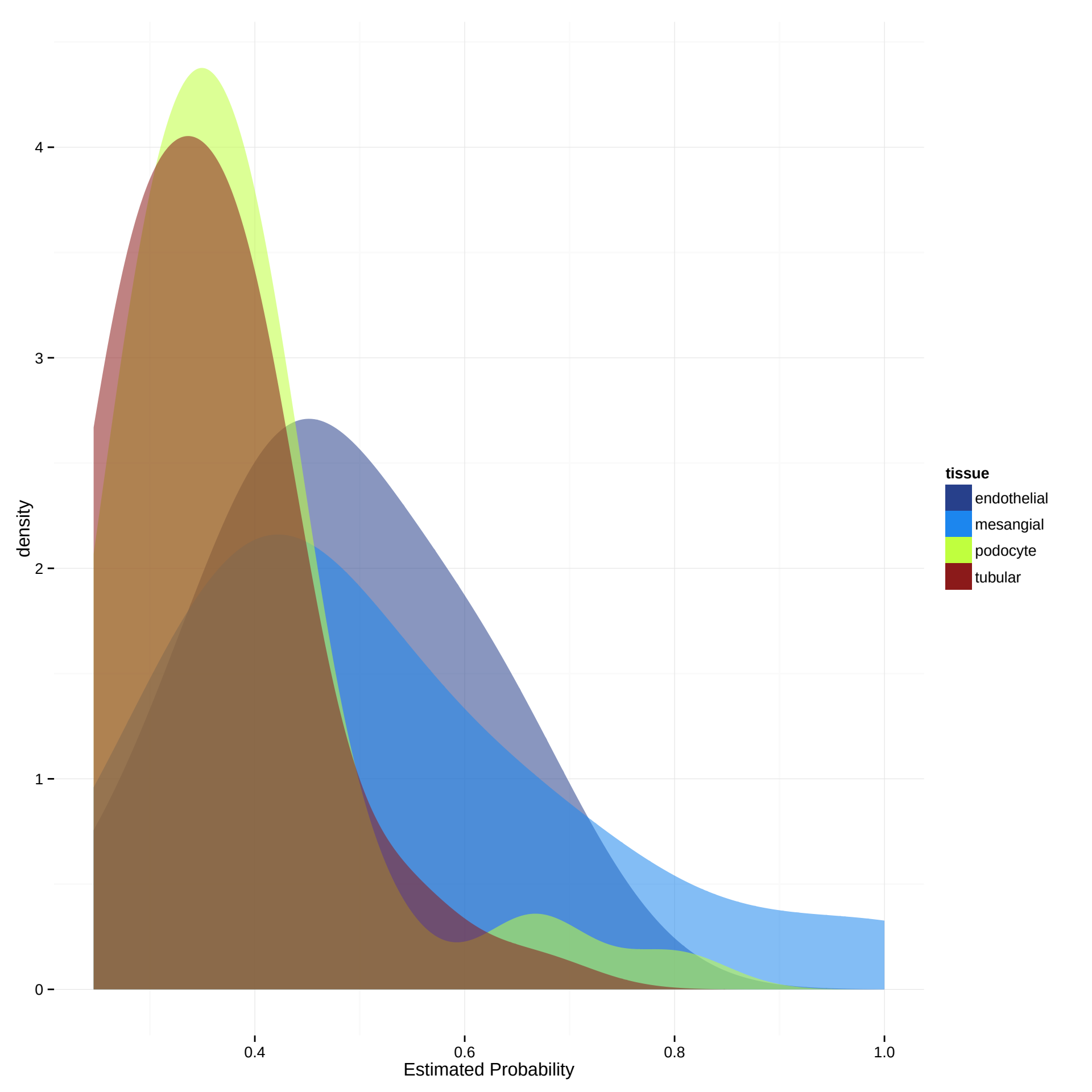



density
8
6
4
2
0
0.4
0.6
0.8
Estimated Probability
tissue
astrocyte
microglia
oligodendro


4
3
2
1
0
density
0.2
0.4
0.6
0.8
Estimated Probability
tissue
fibroblast
keratinocyte
melanocyte

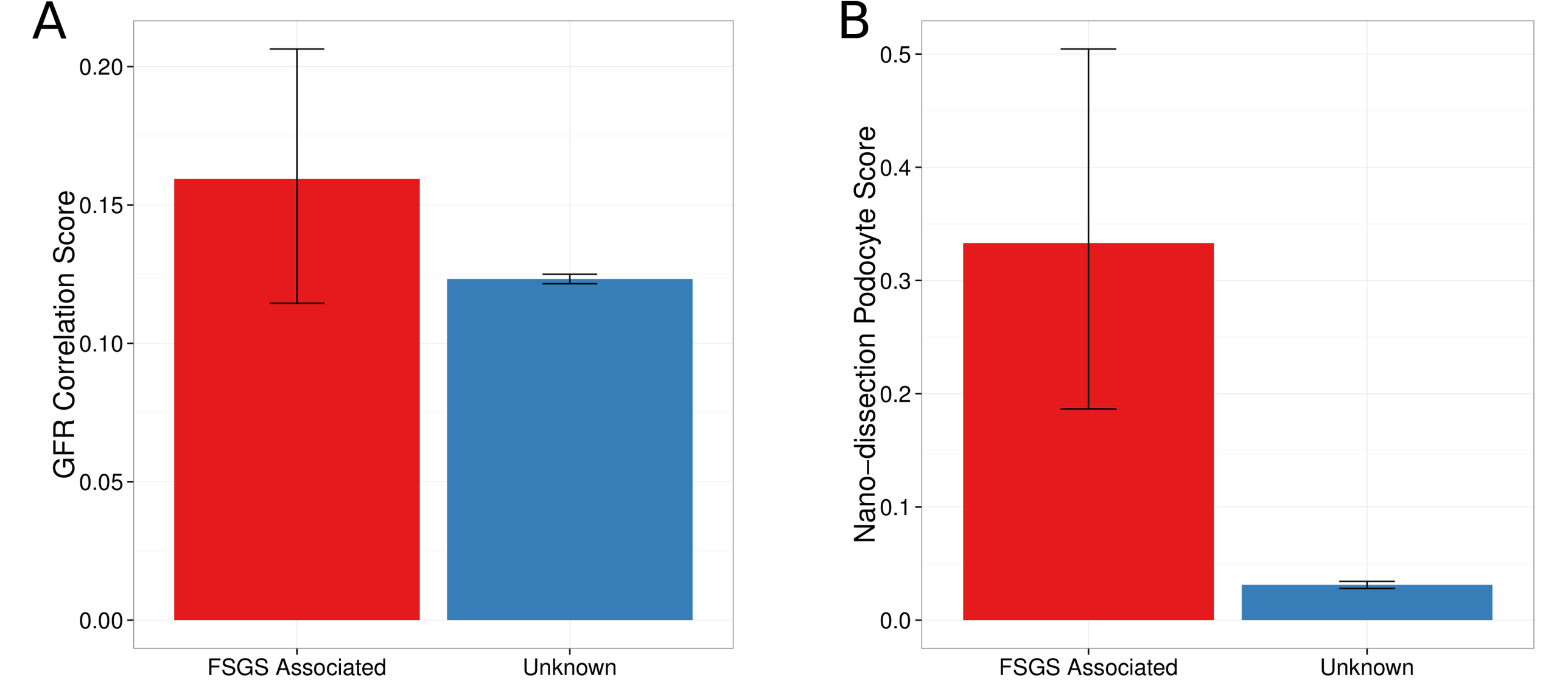

1.0
0.9
0.8
0.7
0.6
0.5
Nano−dissection AUC
5
10
15
20
25
30
35
40
45
Podocyte Genes in Standard


0.85
0.80
0.75
0.70
0.65
0.60
AUC
1
2
3
4
Size of Contaminating Standard (x Size Actual)
Method
Iterative Approach
Linear SVM


density
4
3
2
1
0
0.4
0.6
0.8
1.0
Estimated Probability
tissue
endothelial
mesangial
podocyte
tubular

```
Given: m user supplied standards in ordering of increasing specificity

10.      set all p[k] := 1.0
11.      for k := 1 to m do begin
12.          create k_standard containing all genes from standards k
13.          d[] = results of classification with k_standard
14.          if (lineage-specific validation)
15.              for each negative standard j do begin
16.                  sr_j := \sum_{i=1}^{n_p} R(d_i)
17.                  C_j := n_p n_j + \frac{n_p(n_p+1)}{2} - SR_j
18.                  U_j := \max(C_j, n_p n_j - C_j)
19.                  Z_j := \frac{U_j - \frac{n_j n_p}{2}}{\sqrt{\frac{n_j n_p (n_j + n_p + 1)}{12}}}
20.                  p[k] := p[k] * (1 - cdf(Z_j)) // normal CDF
21.              end;
22.              p[k] := \sqrt[j]{p[k]}
23.          else if (compartment validation)
24.              sr_j := \sum_{i=1}^{n_v} R(|d_i|)
25.              C_j := n_p n_j + \frac{n_p(n_p+1)}{2} - SR_j
26.              U_j := \max(C_j, n_p n_j - C_j)
27.              Z_j := \frac{U_j - \frac{n_j n_p}{2}}{\sqrt{\frac{n_j n_p (n_j + n_p + 1)}{12}}}
28.              p[k] := 1 - cdf(Z_j) // normal CDF
29.          endif
30.      select the kth standard that results in the lowest p[k]
```


A
GFR Correlation Score
0.20
0.15
0.10
0.05
0.00
FSGS Associated
Unknown
B
Nano−dissection Podocyte Score
0.5
0.4
0.3
0.2
0.1
0.0
FSGS Associated
Unknown

Supplementary Table 1. Positive standards for podocytes selected by literature search.

Symbol	Gene Name
ACTN4	actinin, alpha 4
NPHS1	nephrosis 1, congenital, Finnish type (nephrin)
NPHS2	nephrosis 2, idiopathic, steroid-resistant (podocin)
SYNPO	synaptopodin
PTPRO	protein tyrosine phosphatase, receptor type, O
NES	nestin
DDN	dendrin
WT1	Wilms tumor 1
TCF21	transcription factor 21
MAGI1	membrane associated guanylate kinase, WW and PDZ domain containing 1
MAGI2	membrane associated guanylate kinase, WW and PDZ domain containing 2
RAB3A	RAB3A, member RAS oncogene family
KIRREL	kin of IRRE like (Drosophila)
CD2AP	CD2-associated protein
CASK	calcium/calmodulin-dependent serine protein kinase (MAGUK family)
IQGAP1	IQ motif containing GTPase activating protein 1
TJP1	tight junction protein 1 (zona occludens 1)
TRPC6	transient receptor potential cation channel, subfamily C, member 6
DES	desmin
PDPN	podoplanin
PODXL	podocalyxin-like
LMX1B	LIM homeobox transcription factor 1, beta
FAT1	FAT tumor suppressor homolog 1 (Drosophila)
CDH3	cadherin 3, type 1, P-cadherin (placental)
PLCE1	phospholipase C, epsilon 1
CLIC5	chloride intracellular channel 5
CDH13	cadherin 13, H-cadherin (heart)
MME	membrane metallo-endopeptidase
UCHL1	ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)
CD80	CD80 molecule
SV2B	synaptic vesicle glycoprotein 2B
PALLD	palladin, cytoskeletal associated protein
FYN	FYN oncogene related to SRC, FGR, YES
PLAUR	plasminogen activator, urokinase receptor
DAG1	dystroglycan 1 (dystrophin-associated glycoprotein 1)
AGRN	agrin
EFNB1	ephrin-B1
EZR	ezrin
FOXC2	forkhead box C2 (MFH-1, mesenchyme forkhead 1)
MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)
UTRN	utrophin
MYOC	myocilin, trabecular meshwork inducible glucocorticoid response
BASP1	brain abundant, membrane attached signal protein 1
SULF1	sulfatase 1
SCEL	sciellin
CX3CL1	chemokine (C-X3-C motif) ligand 1

Supplementary Table 2. Negative standards for podocytes selected by literature search.

Symbol	Name
ACTA2	actin, alpha 2, smooth muscle, aorta
ADAM10	ADAM metallopeptidase domain 10
AGTR1	angiotensin II receptor, type 1
ALDOB	aldolase B, fructose-bisphosphate
ALOX12	arachidonate 12-lipoxygenase
ANGPT2	angiopoietin 2
APOA4	apolipoprotein A-IV
AQP1	aquaporin 1 (Colton blood group)
AQP11	aquaporin 11
AQP2	aquaporin 2 (collecting duct)
AQP3	aquaporin 3 (Gill blood group)
AQP4	aquaporin 4
AQP6	aquaporin 6, kidney specific
AQP7	aquaporin 7
AQP8	aquaporin 8
BDKRB1	bradykinin receptor B1
BMP7	bone morphogenetic protein 7
CA2	carbonic anhydrase II
CALB1	calbindin 1, 28kDa
CD34	CD34 molecule
CDH16	cadherin 16, KSP-cadherin
CLCNKB	chloride channel Kb
CLDN16	claudin 16
CLDN4	claudin 4
CLDN5	claudin 5
COL6A1	collagen, type VI, alpha 1
CXCL16	chemokine (C-X-C motif) ligand 16
CYP4B1	cytochrome P450, family 4, subfamily B, polypeptide 1
DPYS	dihydropyrimidinase
EDN1	endothelin 1
EHD3	EH-domain containing 3
EMCN	endomucin
ENG	endoglin
ETS1	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)
EXOC3L2	exocyst complex component 3-like 2
F11R	F11 receptor
F2R	coagulation factor II (thrombin) receptor
FBN1	fibrillin 1
FLT1	fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)
FOXC1	forkhead box C1
FXYD2	In multiple Geneids
GAS6	growth arrest-specific 6
GGT1	gamma-glutamyltransferase 1
GLI2	GLI family zinc finger 2
HAVCR1	hepatitis A virus cellular receptor 1
HGD	homogentisate 1,2-dioxygenase
HOXB7	homeobox B7
HPGD	hydroxyprostaglandin dehydrogenase 15-(NAD)

Symbol	Name
HUNK	hormonally up-regulated Neu-associated kinase
ICAM2	intercellular adhesion molecule 2
ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
IGFBP5	insulin-like growth factor binding protein 5
ITGA5	integrin, alpha 5 (fibronectin receptor, alpha polypeptide)
ITGA8	integrin, alpha 8
KDR	kinase insert domain receptor
KHK	ketohexokinase (fructokinase)
KLF6	Kruppel-like factor 6
LDB2	LIM domain binding 2
LMO7	LIM domain 7
LRP2	low density lipoprotein receptor-related protein 2
LTA	lymphotoxin alpha (TNF superfamily, member 1)
MEIS2	In multiple Geneids
MEST	mesoderm specific transcript homolog (mouse)
MPV17L	MPV17 mitochondrial membrane protein-like
PAX2	paired box 2
PAX8	paired box 8
PCDH12	protocadherin 12
PCK2	phosphoenolpyruvate carboxykinase 2 (mitochondrial)
PDGFB	platelet-derived growth factor beta polypeptide
PDGFRB	platelet-derived growth factor receptor, beta polypeptide
PECAM-1	platelet/endothelial cell adhesion molecule
PKD1	polycystic kidney disease 1 (autosomal dominant
PKD2	polycystic kidney disease 2 (autosomal dominant)
PPAP2A	phosphatidic acid phosphatase type 2A
PTGS1	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)
SCNN1G	sodium channel, nonvoltage-gated 1, gamma
SERPINB7	serpin peptidase inhibitor, clade B (ovalbumin), member 7
SERPINE2	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 2
SFRP2	secreted frizzled-related protein 2
SLC12A1	solute carrier family 12 (sodium/potassium/chloride transporters), member 1
SLC12A3	solute carrier family 12 (sodium/chloride transporters), member 3
SLC2A2	solute carrier family 2 (facilitated glucose transporter), member 2
SLC2A9	solute carrier family 2 (facilitated glucose transporter), member 9
SLC34A1	solute carrier family 34 (sodium phosphate), member 1
SLC4A4	solute carrier family 4, sodium bicarbonate cotransporter, member 4
SLC5A2	solute carrier family 5 (sodium/glucose cotransporter), member 2
SLC9A3	solute carrier family 9 (sodium/hydrogen exchanger), member 3
SPP1	secreted phosphoprotein 1
STARD5	In multiple Geneids
TEK	TEK tyrosine kinase, endothelial
THBD	thrombomodulin
THBS1	thrombospondin 1
Thy1	Thy-1 cell surface antigen
TNS1	tensin 1
TNS3	tensin 3
UMOD	uromodulin
VWF	von Willebrand factor

Supplementary Table 3. Top 136 predicted genes by *in silico* nano-dissection approach.

Gene Symbol	Entrez ID	Gene Name
CHI3L1	1116	chitinase 3-like 1 (cartilage glycoprotein-39)
***PCOLCE2*	26577	procollagen C-endopeptidase enhancer 2
***PLA2R1*	22925	phospholipase A2 receptor 1, 180kDa
NDNF	79625	neuron-derived neurotrophic factor
***FGF1*	2246	fibroblast growth factor 1 (acidic)
**CORO2B*	10391	coronin, actin binding protein, 2B
TPPP3	51673	tubulin polymerization-promoting protein family member 3
SPOCK1	6695	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1
CA10	56934	carbonic anhydrase X
DPP6	1804	dipeptidyl-peptidase 6
TNNI1	7135	troponin I type 1 (skeletal, slow)
LRRC2	79442	leucine rich repeat containing 2
TNNT2	7139	troponin T type 2 (cardiac)
ARHGAP19	84986	Rho GTPase activating protein 19
KLK7	5650	kallikrein-related peptidase 7
KLK6	5653	kallikrein-related peptidase 6
TSPAN2	10100	tetraspanin 2
TMEM45A	55076	transmembrane protein 45A
BMP2	650	bone morphogenetic protein 2
ZNF804A	91752	zinc finger protein 804A
**F2R*	2149	coagulation factor II (thrombin) receptor
ANXA1	301	annexin A1
DYNC1I1	1780	dynein, cytoplasmic 1, intermediate chain 1
***NFASC*	23114	neurofascin
TYRO3	7301	TYRO3 protein tyrosine kinase
USP46	64854	ubiquitin specific peptidase 46
**F3*	2152	coagulation factor III (thromboplastin, tissue factor)
PTGDS	5730	prostaglandin D2 synthase 21kDa (brain)
AMIGO2	347902	adhesion molecule with Ig-like domain 2
MCM6	4175	minichromosome maintenance complex component 6
PHYHIP	9796	phytanoyl-CoA 2-hydroxylase interacting protein
MYL9	10398	myosin, light chain 9, regulatory
TNNC1	7134	troponin C type 1 (slow)
MTHFD2	10797	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase
ST3GAL6	10402	ST3 beta-galactoside alpha-2,3-sialyltransferase 6
GPRC5A	9052	G protein-coupled receptor, family C, group 5, member A
**GJA1*	2697	gap junction protein, alpha 1, 43kDa
CDC42EP3	10602	CDC42 effector protein (Rho GTPase binding) 3
**DPYSL3*	1809	dihydropyrimidinase-like 3
**ZNF185*	7739	zinc finger protein 185 (LIM domain)
POLA2	23649	polymerase (DNA directed), alpha 2 (70kD subunit)
MYO1D	4642	myosin ID
RGS10	6001	regulator of G-protein signaling 10
SNCA	6622	synuclein, alpha (non A4 component of amyloid precursor)
SPOCK2	9806	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2
B3GALT2	8707	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 2
NFE2L3	9603	nuclear factor (erythroid-derived 2)-like 3
ZFPM2	23414	zinc finger protein, multitype 2
**SRGAP2*	23380	SLIT-ROBO Rho GTPase activating protein 2
SEMA3G	56920	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3G
PPFIA4	8497	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 4
RASL11B	65997	RAS-like, family 11, member B
ITGA3	3675	integrin, alpha 3 (antigen CD49C, alpha 3 subunit of VLA-3 receptor)
HBEGF	1839	heparin-binding EGF-like growth factor

LOXL1	4016	lysyl oxidase-like 1
ROBO1	6091	roundabout, axon guidance receptor, homolog 1 (Drosophila)
EGR3	1960	early growth response 3
HTRA1	5654	HtrA serine peptidase 1
PLTP	5360	phospholipid transfer protein
FRY	10129	furry homolog (Drosophila)
SST	6750	somatostatin
PPP2R2B	5521	protein phosphatase 2, regulatory subunit B, beta
MYOZ2	51778	myozenin 2
PDE4B	5142	phosphodiesterase 4B, cAMP-specific
PNMA2	10687	paraneoplastic antigen MA2
ITIH5	80760	inter-alpha (globulin) inhibitor H5
***SLK*	9748	STE20-like kinase
ME1	4199	malic enzyme 1, NADP(+)-dependent, cytosolic
****ARHGAP28*	79822	Rho GTPase activating protein 28
PAMR1	25891	peptidase domain containing associated with muscle regeneration 1
****PRKAR2B*	5577	protein kinase, cAMP-dependent, regulatory, type II, beta
CHST2	9435	carbohydrate (N-acetylglucosamine-6-O) sulfotransferase 2
GPNMB	10457	glycoprotein (transmembrane) nmb
***COLGALT2*	23127	collagen beta(1-O)galactosyltransferase 2
MYO1E	4643	myosin IE
PTGER4	5734	prostaglandin E receptor 4 (subtype EP4)
WFS1	7466	Wolfram syndrome 1 (wolframin)
APOD	347	apolipoprotein D
FZD2	2535	frizzled family receptor 2
CCDC91	55297	coiled-coil domain containing 91
HS3ST3A1	9955	heparan sulfate (glucosamine) 3-O-sulfotransferase 3A1
KIAA0040	9674	KIAA0040
***CBLB*	868	Cas-Br-M (murine) ecotropic retroviral transforming sequence b
***ZDHHC6*	64429	zinc finger, DHHC-type containing 6
AIF1	199	allograft inflammatory factor 1
MXRA8	54587	matrix-remodelling associated 8
MFSD6	54842	major facilitator superfamily domain containing 6
H2AFJ	55766	H2A histone family, member J
CDKN1C	1028	cyclin-dependent kinase inhibitor 1C (p57, Kip2)
C1orf21	81563	chromosome 1 open reading frame 21
GMDS	2762	GDP-mannose 4,6-dehydratase
ARHGEF3	50650	Rho guanine nucleotide exchange factor (GEF) 3
KDELR3	11015	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 3
NELL1	4745	NEL-like 1 (chicken)
CACNB2	783	calcium channel, voltage-dependent, beta 2 subunit
RECK	8434	reversion-inducing-cysteine-rich protein with kazal motifs
ELOVL4	6785	ELOVL fatty acid elongase 4
TIMP1	7076	TIMP metallopeptidase inhibitor 1
KBTBD11	9920	kelch repeat and BTB (POZ) domain containing 11
MAFF	23764	v-maf musculoaponeurotic fibrosarcoma oncogene homolog F (avian)
CAND2	23066	cullin-associated and neddylation-dissociated 2 (putative)
S100A8	6279	S100 calcium binding protein A8
COL4A5	1287	collagen, type IV, alpha 5
CTGF	1490	connective tissue growth factor
IGFBP2	3485	insulin-like growth factor binding protein 2, 36kDa
F5	2153	coagulation factor V (proaccelerin, labile factor)
ATP9A	10079	ATPase, class II, type 9A
TUSC3	7991	tumor suppressor candidate 3
UGGT2	55757	UDP-glucose glycoprotein glucosyltransferase 2
TMOD1	7111	tropomodulin 1
EIF1AY	9086	eukaryotic translation initiation factor 1A, Y-linked

PVRL2	5819	poliovirus receptor-related 2 (herpesvirus entry mediator B)
SLAMF8	56833	SLAM family member 8
GNPTAB	79158	N-acetylglucosamine-1-phosphate transferase, alpha and beta subunits
HSPB8	26353	heat shock 22kDa protein 8
PMEPA1	56937	prostate transmembrane protein, androgen induced 1
TCEAL2	140597	transcription elongation factor A (SII)-like 2
LARGE	9215	like-glycosyltransferase
C1S	716	complement component 1, s subcomponent
SCHIP1	29970	schwannomin interacting protein 1
***PDLIM2*	64236	PDZ and LIM domain 2 (mystique)
MYOF	26509	myoferlin
PLP2	5355	proteolipid protein 2 (colonic epithelium-enriched)
IGFBP1	3484	insulin-like growth factor binding protein 1
PIEZO1	9780	piezo-type mechanosensitive ion channel component 1
**NXN*	64359	nucleoredoxin
TMCO3	55002	transmembrane and coiled-coil domains 3
EGR2	1959	early growth response 2
COL4A3	1285	collagen, type IV, alpha 3 (Goodpasture antigen)
HPS5	11234	Hermansky-Pudlak syndrome 5
RNF220	55182	ring finger protein 220
NDN	4692	necdin homolog (mouse)
CMAHP	8418	cytidine monophospho-N-acetylneuraminic acid hydroxylase, pseudogene
AXL	558	AXL receptor tyrosine kinase
HOXA1	3198	homeobox A1
**MPP5*	64398	membrane protein, palmitoylated 5 (MAGUK p55 subfamily member 5)

* denotes genes specifically expressed in podocytes in glomeruli; **denotes genes exclusively expressed in podocytes in whole kidney. This evaluation is based on the IHC images provided by HPA. Genes underlined are also predicted by data analysis using mouse data downloaded from GUDMAP.

Supplementary Table 4. Genes with sufficiently specific staining in HPA to identify localization to renal cell lineages. Bold genes indicate genes that stained only to podocytes within the glomerulus, bold and underlined genes showed podocyte specific staining within the entire kidney.

Gene Symbol	Nano-dissection Score	HPA Identified Lineage	Gene Name
PCOLCE2	**1.08**	**Podocyte in Kidney**	**procollagen C-endopeptidase enhancer 2**
PLA2R1	**1.05**	**Podocyte in Kidney**	**phospholipase A2 receptor 1, 180kDa**
FGF1	**1.00**	**Podocyte in Kidney**	**fibroblast growth factor 1 (acidic)**
CORO2B	**0.95**	**Podocyte in Glomerulus**	**coronin, actin binding protein, 2B**
ZNF804A	0.825	Other Renal	zinc finger protein 804A
F2R	**0.815**	**Podocyte in Glomerulus**	**coagulation factor II (thrombin) receptor**
ANXA1	0.81	Other Renal	annexin A1
NFASC	**0.80**	**Podocyte in Kidney**	**neurofascin**
USP46	0.79	Other Renal	ubiquitin specific peptidase 46
F3	**0.79**	**Podocyte in Glomerulus**	**coagulation factor III (thromboplastin, tissue factor)**
PTGDS	0.78	Other Renal	prostaglandin D2 synthase 21kDa (brain)
GJA1	**0.74**	**Podocyte in Glomerulus**	**gap junction protein, alpha 1, 43kDa**
DPYSL3	**0.73**	**Podocyte in Glomerulus**	**dihydropyrimidinase-like 3**
ZNF185	**0.73**	**Podocyte in Glomerulus**	**zinc finger protein 185 (LIM domain)**
SRGAP2	**0.70**	**Podocyte in Glomerulus**	**SLIT-ROBO Rho GTPase activating protein 2**
SEMA3G	0.70	Other Renal	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3G
SLK	**0.65**	**Podocyte in Glomerulus**	**STE20-like kinase**
ARHGAP28	**0.65**	**Podocyte in Kidney**	**Rho GTPase activating protein 28**
PRKAR2B	**0.64**	**Podocyte in Kidney**	**protein kinase, cAMP-dependent, regulatory, type II, beta**
GPNMB	0.63	Other Renal	glycoprotein (transmembrane) nmb
COLGALT2	**0.63**	**Podocyte in Glomerulus**	**collagen beta (1-O) galactosyltransferase 2**
PTGER4	0.63	Other Renal	prostaglandin E receptor 4 (subtype EP4)
CBLB	**0.61**	**Podocyte in Glomerulus**	**Cas-Br-M (murine) ecotropic retroviral transforming sequence b**
ZDHHC6	**0.61**	**Podocyte in Glomerulus**	**zinc finger, DHHC-type containing 6**
MFSD6	0.61	Other Renal	major facilitator superfamily domain containing 6
S100A8	0.58	Other Renal	S100 calcium binding protein A8
PVRL2	0.57	Other Renal	poliovirus receptor-related 2 (herpesvirus entry mediator B)
LARGE	0.56	Other Renal	like-glycosyltransferase
PDLIM2	**0.55**	**Podocyte in Kidney**	**PDZ and LIM domain 2 (mystique)**
NXN	**0.55**	**Podocyte in Glomerulus**	**nucleoredoxin**
MPP5	**0.53**	**Podocyte in Glomerulus**	**membrane protein, palmitoylated 5 (MAGUK p55 subfamily member 5)**

Supplementary Table 5. GUDMAP gene expression datasets used in this study.

GUDMAP ID	GEO Sample ID	Sample ID	Sample Description
12336	GSM429038	Adult_MafB_GFP1	Visceral epithelium (syn: podocytes layer)
12337	GSM429039	Adult_MafB_GFP2	Visceral epithelium (syn: podocytes layer)
12338	GSM429040	Adult_MafB_GFP3	Visceral epithelium (syn: podocytes layer)
13521	GSM519446	Adult_Meis_GFP1	Glomerular mesangial
13522	GSM519447	Adult_Meis_GFP2	Glomerular mesangial
13523	GSM519448	Adult_Meis_GFP3	Glomerular mesangial
13694	GSM559844	Tie2_Glom1_GFP1	Glomerular Capillary System
13695	GSM559845	Tie2_Glom1_GFP2	Glomerular Capillary System
13696	GSM559846	Tie2_Glom1_GFP3	Glomerular Capillary System

Supplementary Table 6. 102 highly expressed genes in podocytes in comparison with the other two major cell types in mouse glomerulus. Analysis is based on array data from GUDMAP (see website).

Mus musculus gene_ID	Mus musculus gene_symbol	Homolo-Gene-ID	Homo sapiens gene_ID	Homo sapiens gene_symbol	podo vs endo (LogFC)	podo vs endo (FC)	podo vs mes (logFC)	podo vs mes (FC)
13482	*Dpp4*	3279	1803	*DPP4*	6.26	76.38	6.59	96.59
18479	***Pak1*	1936	5058	***PAK1*	5.75	53.87	5.98	63.01
13809	***Enpep*	68216	2028	***ENPEP*	6.26	76.49	5.37	41.38
15229	*Foxd1*	3290	2297	*FOXD1*	5.66	50.60	5.65	50.14
14181	*Fgfbp1*	3766	9982	*FGFBP1*	5.57	47.42	5.68	51.44
103142	*Rdh9*	74505	8608	*RDH16*	5.68	51.30	5.38	41.58
15478	*Hs3st3a1*	88735	9955	*HS3ST3A1*	5.46	44.04	5.48	44.78
69908	*Rab3b*	2149	5865	*RAB3B*	5.17	35.91	5.16	35.68
269181	*Mgat4a*	8153	11320	*MGAT4A*	5.06	33.47	4.96	31.23
66260	*Tmem54*	11937	113452	*TMEM54*	4.87	29.24	5.02	32.55
320910	*Itgb8*	18567	3696	*ITGB8*	4.81	27.98	4.86	29.00
71934	*Car13*	75207	377677	*CA13*	4.84	28.66	4.82	28.31
241447	*Cers6*	72228	253782	*CERS6*	4.81	27.98	4.81	28.09
17930	*Myom2*	2953	9172	*MYOM2*	4.75	26.96	4.86	29.06
20349	*Sema3e*	8247	9723	*SEMA3E*	4.72	26.38	4.74	26.68
78910	*Asb15*	43797	142685	*ASB15*	4.62	24.51	4.72	26.42
12577	*Cdkn1c*	58	1028	*CDKN1C*	4.69	25.82	4.64	25.00
544963	****Iqgap2*	101543	10788	****IQGAP2*	4.70	26.08	4.31	19.86
210530	*Leprel1*	10062	55214	*LEPREL1*	6.43	85.95	2.59	6.01
54366	*Ctnnal1*	2815	8727	*CTNNAL1*	5.05	33.24	3.94	15.30
226352	*Epb4.1l5*	32492	57669	*EPB41L5*	5.02	32.42	3.66	12.66
212531	*Sh3bgrl2*	12876	83699	*SH3BGRL2*	4.34	20.31	4.26	19.19
231532	*Arhgap24*	32754	83478	*ARHGAP24*	4.86	29.05	3.73	13.26
54670	*Atp8b1*	21151	5205	*ATP8B1*	3.66	12.65	4.92	30.31
268970	****Arhgap28*	18264	79822	****ARHGAP28*	4.78	27.38	3.79	13.83
109222	*Rarres1*	2166	5918	*RARRES1*	4.24	18.85	4.32	20.03
11535	*Adm*	873	133	*ADM*	4.65	25.18	3.84	14.33
69354	*Slc38a4*	75077	55089	*SLC38A4*	4.04	16.43	4.27	19.31
226101	*Myof*	40882	26509	*MYOF*	4.36	20.58	3.58	11.93
72960	*Top1mt*	43082	116447	*TOP1MT*	3.72	13.15	4.19	18.23
117600	*Srgap1*	56898	57522	*SRGAP1*	3.49	11.27	4.41	21.22

68813	*Dock5*	57016	80005	*DOCK5*	4.00	16.04	3.82	14.14
214345	*Lrrc1*	44445	55227	*LRRC1*	3.82	14.11	3.74	13.35
14400	*Gabrb1*	20221	2560	*GABRB1*	3.65	12.52	3.91	15.02
70415	*2610018G03Rik*	84402	51765	*MST4*	3.66	12.63	3.83	14.25
69454	*Clic3*	3432	9022	*CLIC3*	3.67	12.76	3.80	13.94
208890	*Slc26a7*	13770	115111	*SLC26A7*	3.64	12.49	3.75	13.43
195208	*Dcdc2a*	9483	51473	*DCDC2*	3.59	12.08	3.79	13.81
26878	*B3galt2*	74512	8707	*B3GALT2*	3.76	13.53	3.61	12.21
140792	*Colec12*	34248	81035	*COLEC12*	3.58	11.99	3.72	13.13
170743	*Tlr7*	75060	51284	*TLR7*	3.68	12.85	3.55	11.70
13134	**Dach1*	7288	1602	***DACH1*	3.97	15.66	3.22	9.29
56318	*Acpp*	55552	55	*ACPP*	3.56	11.82	3.61	12.19
14397	*Gabra4*	631	2557	*GABRA4*	3.46	11.03	3.68	12.85
109323	*C1qtnf7*	12900	114905	*C1QTNF7*	3.56	11.81	3.52	11.48
54381	*Cpq*	9401	10404	*CPQ*	4.77	27.22	2.26	4.78
333883	*Cd59b*	56386	966	*CD59*	3.42	10.68	3.59	12.04
13197	*Gadd45a*	1449	1647	*GADD45A*	3.59	12.07	3.39	10.49
68178	***Cgnl1*	41901	84952	***CGNL1*	3.93	15.27	3.05	8.28
170772	*Glcci1*	15773	113263	*GLCCI1*	3.57	11.86	3.40	10.59
72607	*Usp13*	68372	8975	*USP13*	3.39	10.45	3.58	11.93
330267	***Thsd7a*	46582	221981	***THSD7A*	2.91	7.54	4.05	16.52
217265	*Abca5*	10263	23461	*ABCA5*	3.76	13.56	3.14	8.79
107515	*Lgr4*	10226	55366	*LGR4*	3.74	13.34	3.10	8.57
16173	*Il18*	1200	3606	*IL18*	3.15	8.89	3.44	10.83
81907	*Tmem108*	18688	66000	*TMEM108*	2.93	7.63	3.63	12.41
107895	*Mgat5*	1808	4249	*MGAT5*	3.30	9.83	3.24	9.46
231503	*Tmem150c*	18818	441027	*TMEM150C*	3.21	9.25	3.31	9.90
73094	*Sgip1*	13001	84251	*SGIP1*	3.46	11.01	3.05	8.27
30953	*Schip1*	32200	29970	*SCHIP1*	2.68	6.40	3.80	13.97
76459	*Car12*	20327	771	*CA12*	3.21	9.24	3.26	9.56
56312	*Nupr1*	8229	26471	*NUPR1*	3.83	14.20	2.63	6.21
242122	*Vtcn1*	11627	79679	*VTCN1*	2.97	7.86	3.41	10.61
68024	*Hist1h2bc*	113763	8970	*HIST1H2BJ*	2.91	7.54	3.38	10.42
70747	*Tspan2*	21169	10100	*TSPAN2*	2.60	6.05	3.62	12.32
108897	*Aif1l*	12863	83543	*AIF1L*	2.87	7.31	3.34	10.16
14420	*Galc*	124	2581	*GALC*	3.82	14.16	2.39	5.24

107351	*Kank1*	17706	23189	*KANK1*	3.62	12.29	2.59	6.03
20971	*Sdc4*	31121	6385	*SDC4*	3.40	10.58	2.77	6.84
114249	*Npnt*	14138	255743	*NPNT*	2.56	5.88	3.55	11.68
218518	*Marveld2*	27037	153562	*MARVELD2*	2.91	7.51	3.14	8.83
97114	*Hist2h3c2*	110949	8350	*HIST1H3A*	2.71	6.54	3.32	9.96
74051	*Steap2*	17682	261729	*STEAP2*	3.02	8.11	2.96	7.76
13645	*Egf*	1483	1950	*EGF*	3.20	9.20	2.73	6.63
22339	*Vegfa*	2534	7422	*VEGFA*	2.24	4.71	3.69	12.86
102545	*Cmtm7*	15882	112616	*CMTM7*	2.72	6.61	3.19	9.10
19271	*Ptprj*	2130	5795	*PTPRJ*	2.68	6.39	3.15	8.85
338365	*Slc41a2*	9870	84102	*SLC41A2*	3.02	8.11	2.76	6.75
12955	*Cryab*	68209	1410	*CRYAB*	3.43	10.76	2.35	5.08
140580	*Elmo1*	56685	9844	*ELMO1*	2.57	5.96	3.18	9.09
239083	*Ccnb1ip1*	18819	57820	*CCNB1IP1*	2.83	7.12	2.92	7.55
66848	*Fuca2*	119	2519	*FUCA2*	3.18	9.06	2.52	5.75
211949	*Spsb4*	25315	92369	*SPSB4*	2.73	6.62	2.94	7.66
54003	*Nell2*	4488	4753	*NELL2*	2.72	6.59	2.94	7.69
67451	*Pkp2*	3364	5318	*PKP2*	2.88	7.39	2.68	6.41
353169	*Slc2a12*	59263	154091	*SLC2A12*	2.71	6.53	2.85	7.22
110310	*Krt7*	4058	3855	*KRT7*	2.80	6.96	2.72	6.61
210029	*Metrnl*	16993	284207	*METRNL*	2.93	7.62	2.43	5.40
12828	*Col4a3*	68033	1285	*COL4A3*	2.75	6.71	2.58	5.98
16419	*Itgb5*	20511	3693	*ITGB5*	2.96	7.79	2.31	4.97
18578	*Pde4b*	1953	5142	*PDE4B*	2.50	5.66	2.76	6.78
100039795	*Ildr2*	52388	387597	*ILDR2*	2.54	5.83	2.70	6.50
72634	*Tdrkh*	4999	11022	*TDRKH*	2.41	5.33	2.66	6.34
244141	*Nars2*	57002	79731	*NARS2*	2.57	5.92	2.51	5.68
320685	*Dctd*	55618	1635	*DCTD*	2.45	5.47	2.62	6.13
71648	*Optn*	11085	10133	*OPTN*	2.50	5.65	2.41	5.32
12829	*Col4a4*	20071	1286	*COL4A4*	2.46	5.49	2.38	5.21
22094	*Tshb*	463	7252	*TSHB*	2.28	4.84	2.56	5.90
207965	*Mettl21d*	56995	79609	*METTL21D*	2.45	5.48	2.38	5.20
16398	*Itga2*	1662	3673	*ITGA2*	2.34	5.05	2.46	5.51
76969	*Chst1*	2711	8534	*CHST1*	2.28	4.86	2.45	5.47
11443	*Chrnb1*	594	1140	*CHRNB1*	2.22	4.66	2.28	4.85

Abbreviations are: Podo, podocytes (visceral epithelial cells); Endo, glomerular endothelial cells; Mes, glomerular mesangial cells. * denotes genes specifically expressed in podocytes in glomeruli; **denotes genes exclusively expressed in podocytes in whole kidney. This evaluation is based on the IHC images provided by HPA. Genes underlined are also predicted by *in silico* nano-dissection.

Supplementary Table 7. Demographics of 139 patients. Age and eGFR are presented as means ± standard deviation. eGFR was calculated using the MDRD equation.

Disease type	CKD patients (n=139)	eGFR (mL/min per 1.73 m2)	Age (years)	Gender (male/female)
LN	30	63.7±29.4	34.7±13.4	7m/23f
IgAN	26	73.9±37.0	36.4±14.1	19m/7f
MGN	21	84.8±40.1	53.7±18.0	12m/9f
FSGS	17	70.9±37.0	44.6±15.8	9m/8f
HTN	15	40.9±23.8	57.1±11.8	12m/3f
DN	12	52.9±30.3	54.8±10.9	8m/4f
MCD	12	97.3±45.7	37.4±17.7	7m/5f
FSGS-MCD	6	85.7±22.0	47.9±15.7	2m/4f