Supplemental Material – Table of Contents

C. Supplemental Tables

Supplemental Table 1. Anonymized IDs for original NEPTUNE, ERCB, expanded NEPTUNE, validation, and high-proteinuria participants included in study, and PLA2R statuses for NEPTUNE participants.

Supplemental Table 2. Baseline clinical characteristics of NEPTUNE, ERCB, and validation cohort participants included in the study, tabulated by diagnosis as well as by MN vs. all other diagnoses.

Supplemental Table 3. Mean AUCs for disease-specific classifiers.

Supplemental Table 4. MN-specific gene list and overlap with genes previously reported to be differentially expressed in a Heymann nephritis rat model.

Supplemental Table 5. Module enrichments within the MN-specific gene list.

Supplemental Table 6. MN-specific genes overlapping with podocyte markers from single-cell and single-nucleus RNA sequencing studies.

Supplemental Table 7. Module enrichments within the podocyte-expressed MN-specific gene list.

Supplemental Table 8. Comparison of expression of canonical podocyte marker genes, eigengenes, and GFRs for participants with MN vs. all other diagnoses and participants with MN vs. participants with MCD.

Supplemental Table 9: Metadata for the Human Protein Atlas images used to generate Supplemental Figure 12.

A. Supplemental Methods

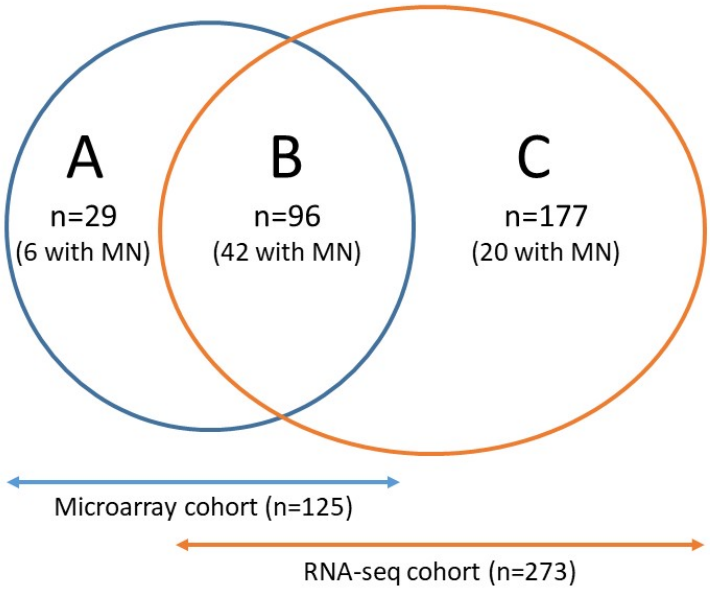
Generation of the Human Protein Atlas (HPA) image composite depicting glomerular protein expression of top MN classifier genes:

A single author (LHB) queried the tissue atlas (kidney) component of the Human Protein Atlas (www.proteinatlas.org) for the top 25 MN classifier genes. Images were collected for each antibody reactive with the protein of interest. In most cases, a representative image from each of 3 separate tissue sections was available for each antibody that had been generated to the protein of interest). For each available HPA kidney tissue image, a representative glomerulus was digitally enlarged at the website to the point that the scale bar read 20 microns. A screenshot was taken and converted to a JPEG file using Microsoft Windows Paint version 6.1. The screenshot included the glomerulus and the information box with the antibody catalog number, and the age and gender of the case from which the tissue was derived. The JPEG was inserted into a Microsoft Office PowerPoint 2013 file and the specific URL was saved with the image.

For two of the top 25 MN classifier genes (SARAF and BMP2) there were no kidney images available at the HPA site. All the other collected images were independently scored by two other authors (G.L. and N.H.) for the presence of glomerular and podocyte staining. The results were compiled and any discrepancies were adjudicated by LHB. For each classifier gene/protein, the most representative image that conveyed the presence or absence of podocyte staining was moved to another PowerPoint file for generation of the composite image shown as Supplemental Figure 12. The age and gender associated with the tissue section, as well as the antibody's HPA identifier and URL from which the image was taken are listed were saved as a separate Excel file (Supplemental Table 9).

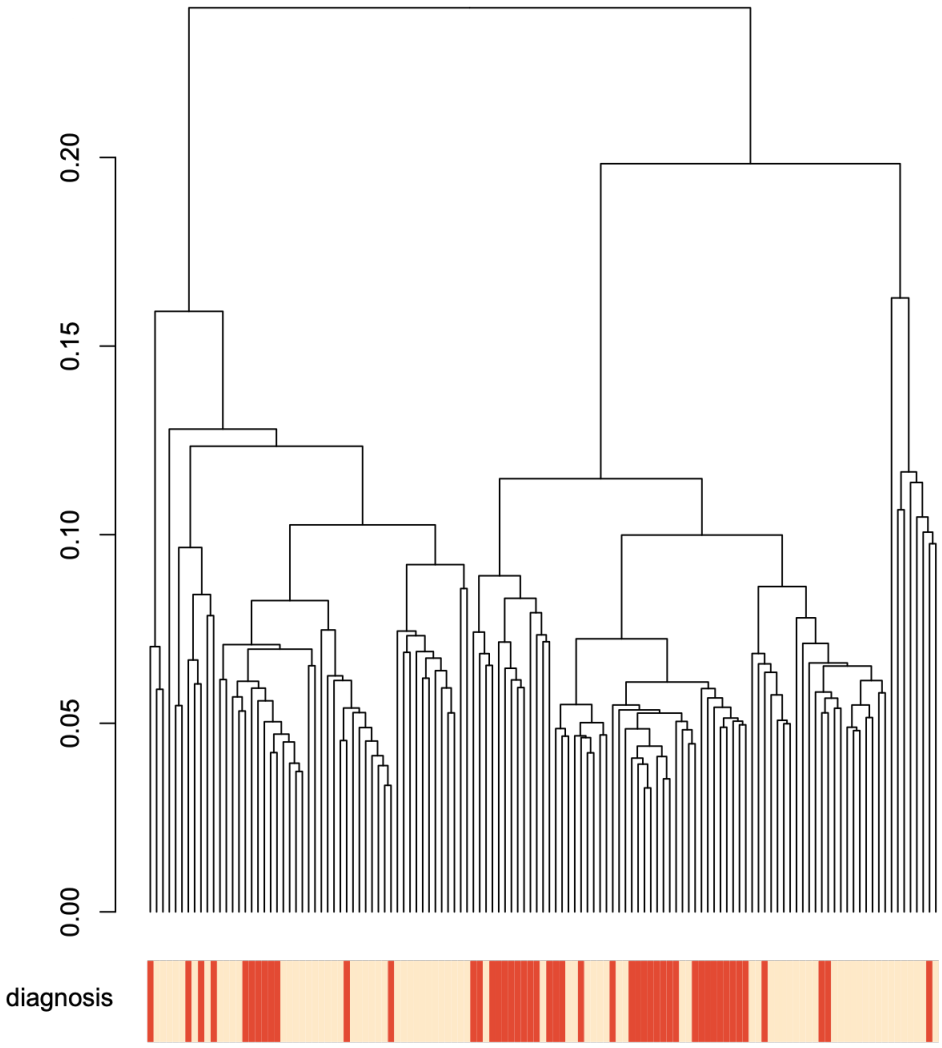
B. Supplemental Figures

Supplemental Figure 1. Schematic of original, validation, and expanded cohorts. The blue circle indicates participants with microarray data (original cohort; 125 participants including 48 with MN) and the orange circle indicates participants with RNA-seq data (expanded cohort; 273 participants including 62 with MN). Region A in the Venn diagram indicates the set of participants in the original cohort with only microarray data available (29 participants, including 6 with MN). Region B includes participants from both the original (microarray) and expanded (RNA-seq) cohorts all of whom have RNA-seq data available (96 participants, including 42 with MN). Region C indicates participants in the validation cohort who only have RNA-seq data and are not present in the original cohort (177 participants, including 20 with MN). The original cohort includes all participants with microarray data (regions A + B); the expanded cohort includes all participants with RNA-seq data (regions B + C); the validation cohort includes participants with only RNA-seq data (region C).

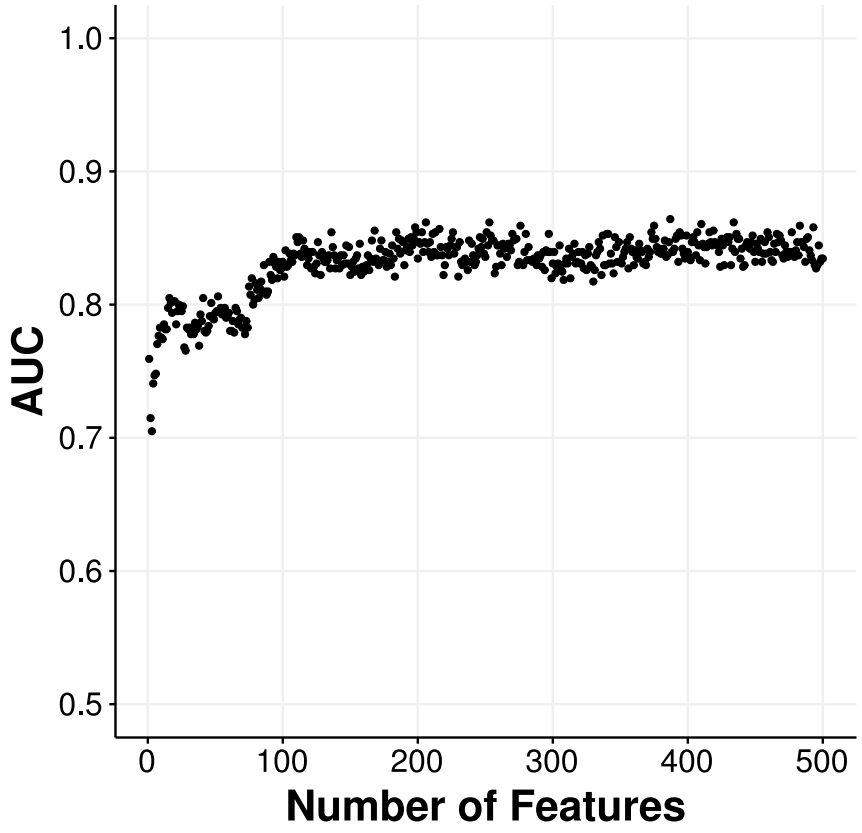


Original cohort = all participants with glomerular microarray data = **A** + **B** (n=125, 48 with MN)
Expanded cohort = all participants with glomerular RNA-seq data = **B** + **C** (n=273, 62 with MN)
Validation cohort = all new participants with glomerular RNA-seq data = **C** (n=177, 20 with MN)

Supplemental Figure 2. Hierarchical clustering of subjects by diagnosis using all genes. Clustering was performed using the swamp R package. All genes were used as features and the samples were colored by patient diagnosis. Red, MN; beige, diagnoses other than MN.

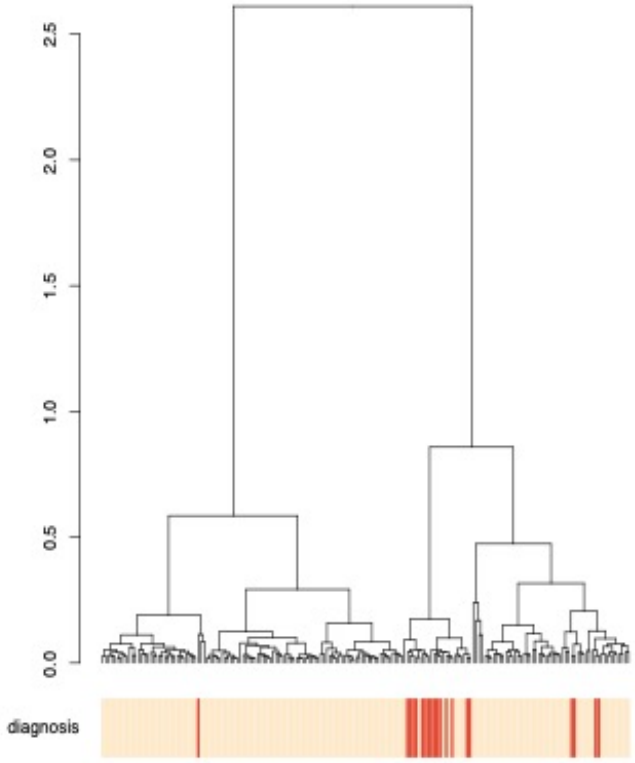


Supplemental Figure 3. Performance of MN classifier versus number of features used to train the classifier. The classifier distinguishing MN subjects from subjects with other glomerulonephropathies has similar performance when trained on top ~100 features as when including additional features. A random forest classifier was trained on two-thirds of the participants to distinguish MN subjects from subjects with other glomerulonephropathies using all genes as features. Genes were then sorted based on their Gini importance in this classifier. A random forest classifier was then trained using the top k genes using the same subset of subjects as previously and evaluated on the remaining one-third of the participants, for k ranging from 1 to 500. Performance continued to increase until ~100 features were added and remained stable thereafter. Random forest classification was performed using the sklearn python library.

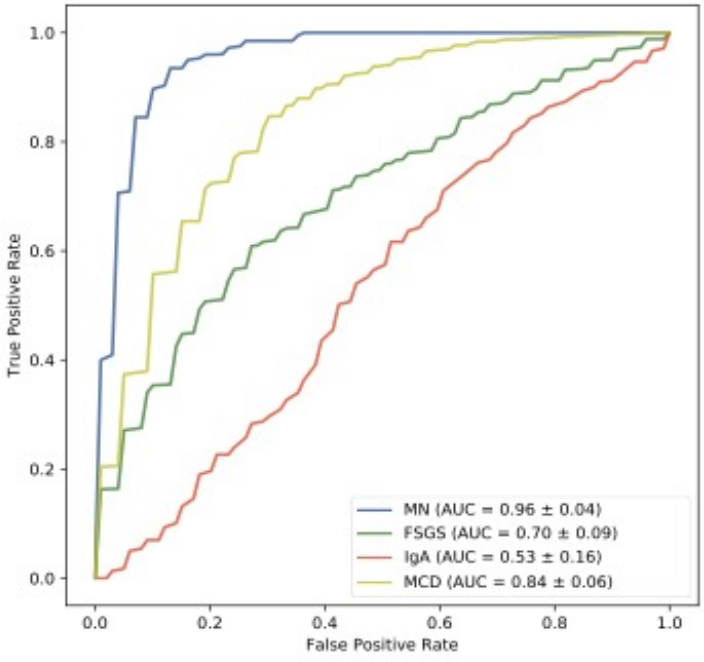


Supplemental Figure 4. A glomerular gene expression signature can classify MN relative to other causes of the nephrotic syndrome in the validation cohort. (A) MN participants cluster by expression of top glomerular compartment genes that are predictive of diagnosis in the validation cohort. First, we fit a multi-class random forest classifier with 500 estimators and a maximum of 1000 features to predict diagnosis from glomerular gene expression in the validation cohort using the Python sklearn package. The top 500 genes with highest Gini importance for this classification task were then selected. We then clustered all patients in the validation cohort based on the expression of these selected top genes using the swamp R package, and colored the samples by patient diagnosis (red, MN; beige, diagnoses other than MN). (B) MN has the highest prediction accuracy (AUC) across diseases in the validation cohort, based on a random forest classifier with five-fold cross-validation. For each disease, we train a random forest classifier with 500 estimators and a maximum of 1000 features to distinguish patients with that diagnosis from all other diagnoses. AUCs are averaged across ten classification runs for each diagnosis.

A

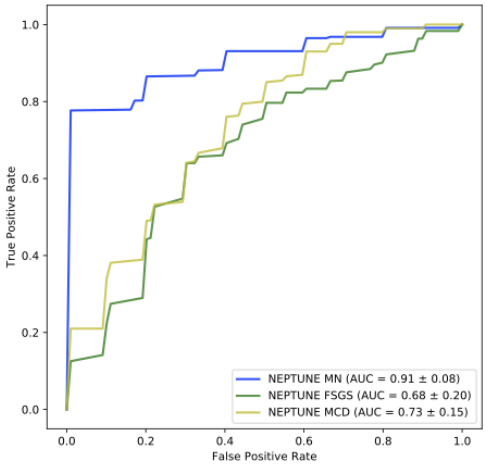

2.5
2.0
1.5
1.0
0.5
0.0
diagnosis

B

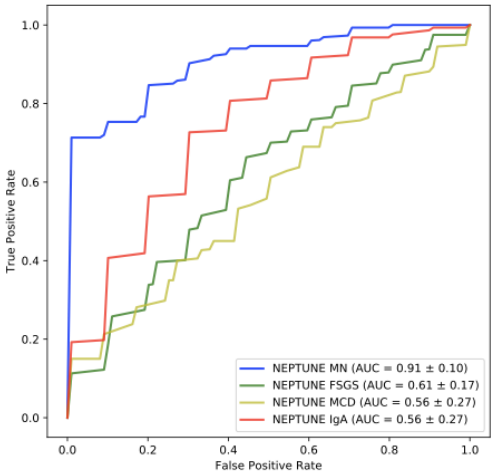

True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
MN (AUC = 0.96 ± 0.04)
FSGS (AUC = 0.70 ± 0.09)
IgA (AUC = 0.53 ± 0.16)
MCD (AUC = 0.84 ± 0.06)

Supplemental Figure 5. Prediction accuracy for distinguishing proteinuric participants with MN from participants with all other diagnoses. (A) High-proteinuria participants in original NEPTUNE cohort. (B) Low-proteinuria participants in original NEPTUNE cohort. (C) High-proteinuria participants in expanded NEPTUNE cohort. (D) Low-proteinuria participants in expanded NEPTUNE cohort. (E) High-proteinuria participants in ERCB cohort. (F) Low-proteinuria participants in ERCB cohort. Classification accuracy for IgA nephropathy is not shown for high-proteinuria cohorts due to an insufficient number of proteinuric participants with IgA nephropathy in these cohorts. Classification accuracy for MCD is not shown for the low-proteinuria ERCB cohort due to an insufficient number of non-proteinuric participants with MCD in this cohort. The MN curve is colored blue in all plots.

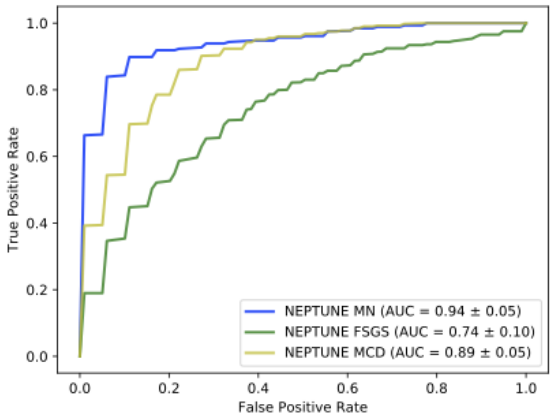
A


True Positive Rate
False Positive Rate
NEPTUNE MN (AUC = 0.91 ± 0.08)
NEPTUNE FSGS (AUC = 0.68 ± 0.20)
NEPTUNE MCD (AUC = 0.73 ± 0.15)

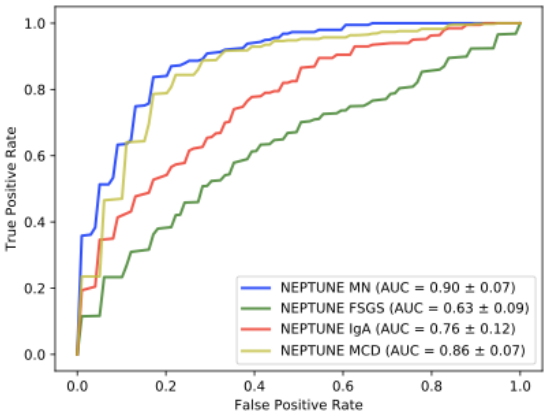
B


True Positive Rate
False Positive Rate
NEPTUNE MN (AUC = 0.91 ± 0.10)
NEPTUNE FSGS (AUC = 0.61 ± 0.17)
NEPTUNE MCD (AUC = 0.56 ± 0.27)
NEPTUNE IgA (AUC = 0.56 ± 0.27)

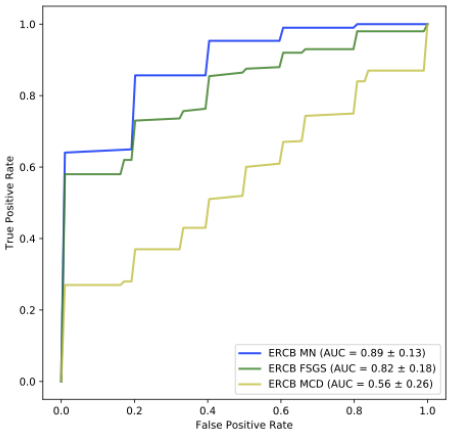
C


True Positive Rate
False Positive Rate
NEPTUNE MN (AUC = 0.94 ± 0.05)
NEPTUNE FSGS (AUC = 0.74 ± 0.10)
NEPTUNE MCD (AUC = 0.89 ± 0.05)

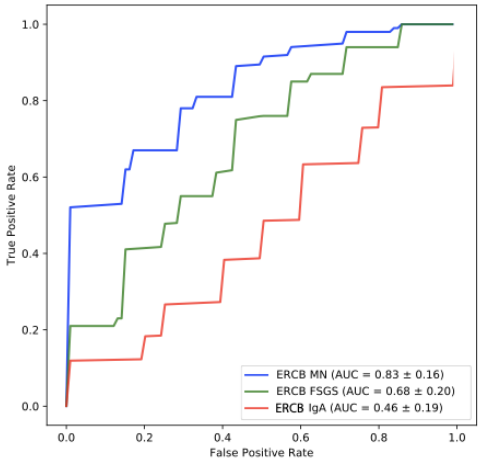
D


True Positive Rate
False Positive Rate
NEPTUNE MN (AUC = 0.90 ± 0.07)
NEPTUNE FSGS (AUC = 0.63 ± 0.09)
NEPTUNE IgA (AUC = 0.76 ± 0.12)
NEPTUNE MCD (AUC = 0.86 ± 0.07)

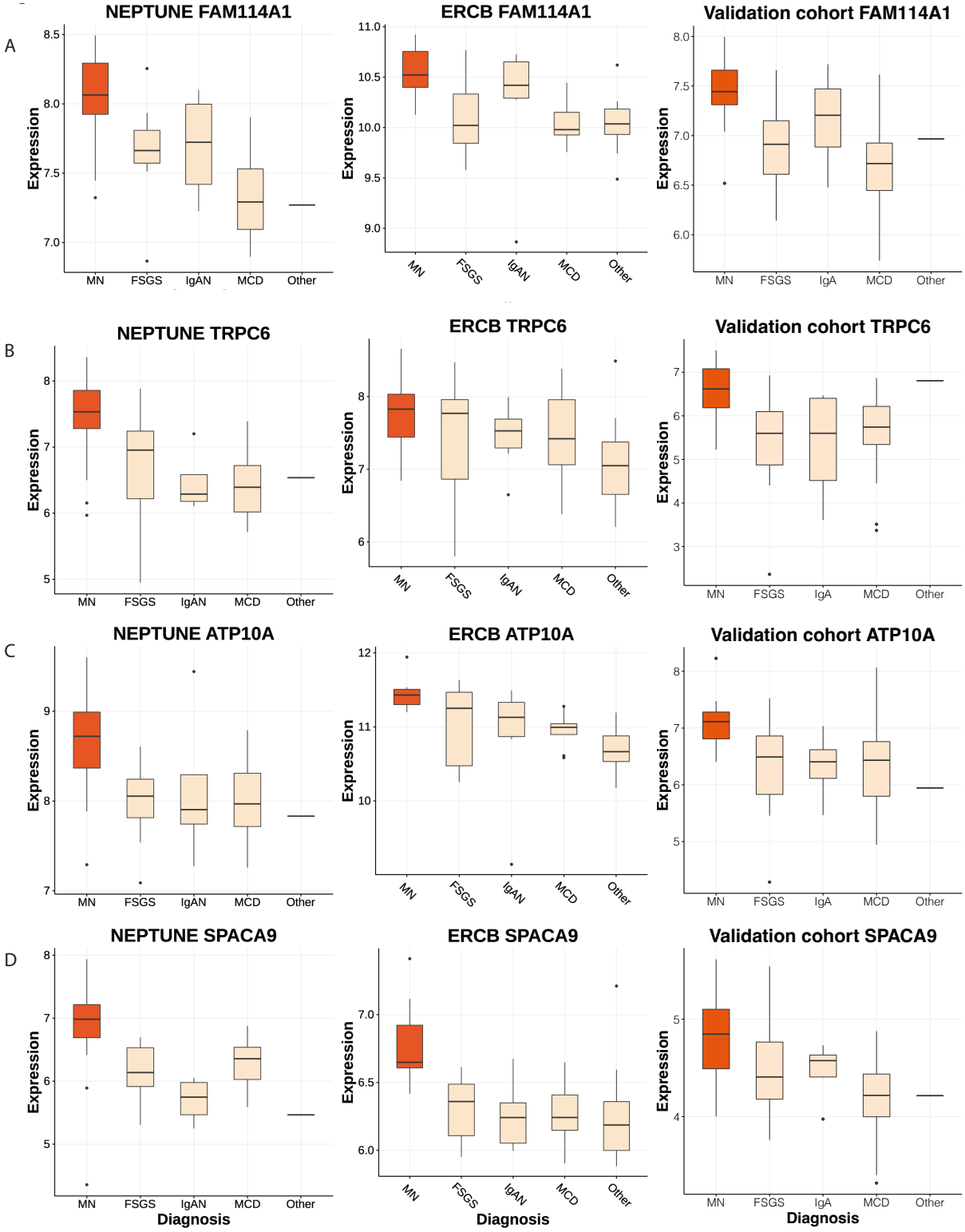
E


True Positive Rate
False Positive Rate
ERCB MN (AUC = 0.89 ± 0.13)
ERCB FSGS (AUC = 0.82 ± 0.18)
ERCB MCD (AUC = 0.56 ± 0.26)

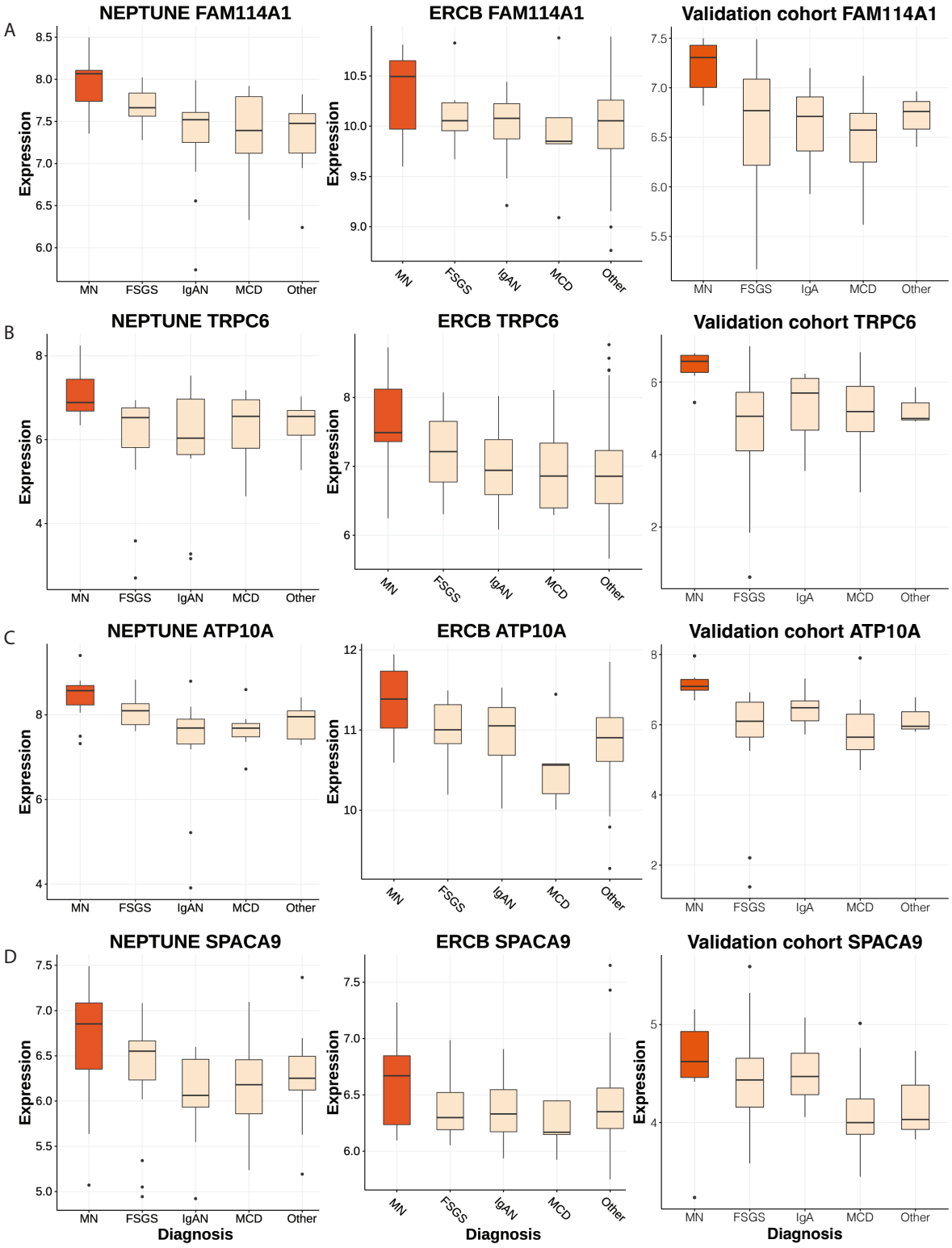
F


True Positive Rate
False Positive Rate
ERCB MN (AUC = 0.83 ± 0.16)
ERCB FSGS (AUC = 0.68 ± 0.20)
ERCB IgA (AUC = 0.46 ± 0.19)

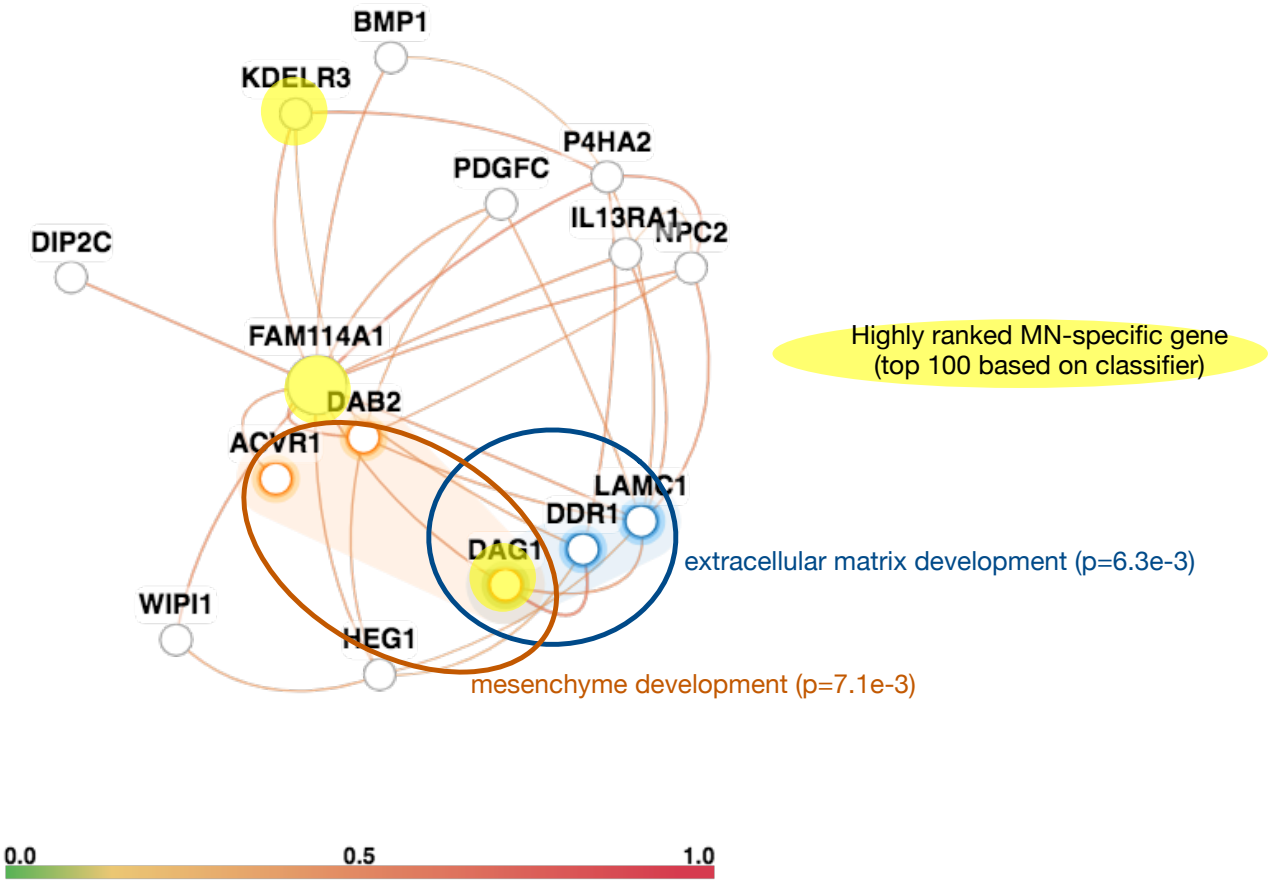
Supplemental Figure 6. Top genes implicated in MN based on analysis of NEPTUNE cohort are also upregulated in the high-proteinuria subsets of original NEPTUNE, ERCB, and validation cohorts.



Supplemental Figure 7. Top genes implicated in MN based on analysis of NEPTUNE cohort are also upregulated in the low-proteinuria subsets of original NEPTUNE, ERCB, and validation cohorts.



Supplemental Figure 8. Neighbors of FAM114A1 in the kidney functional network include other highly ranked MN-specific genes (yellow highlight) as well as genes enriched in extracellular matrix development (blue) and mesenchyme development (orange) processes. FAM114A1 was queried in the kidney-specific functional network at https://hb.flatironinstitute.org. The kidney-specific functional network was generated by integrating thousands of publicly available experimental datasets to predict how likely it is that each pair of genes functions in the same process in kidney tissue. The enriched processes for the top neighbors of FAM114A1 are highlighted in the figure.

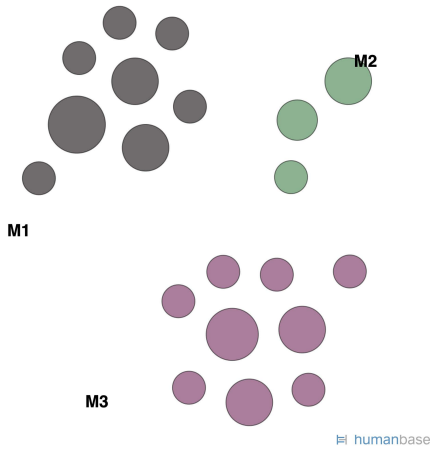


Supplemental Figure 9. Gene ontology enrichments and functional modules for podocyte-expressed MN-specific genes. (A) Gene ontology enrichments for 65 MN genes that are specifically expressed in podocytes (Supplemental Table 5) are identified using the GOLEM gene ontology tool. (B) Functional modules are identified using community clustering in the kidney-specific functional network for podocyte-expressed MN-specific genes.

A.

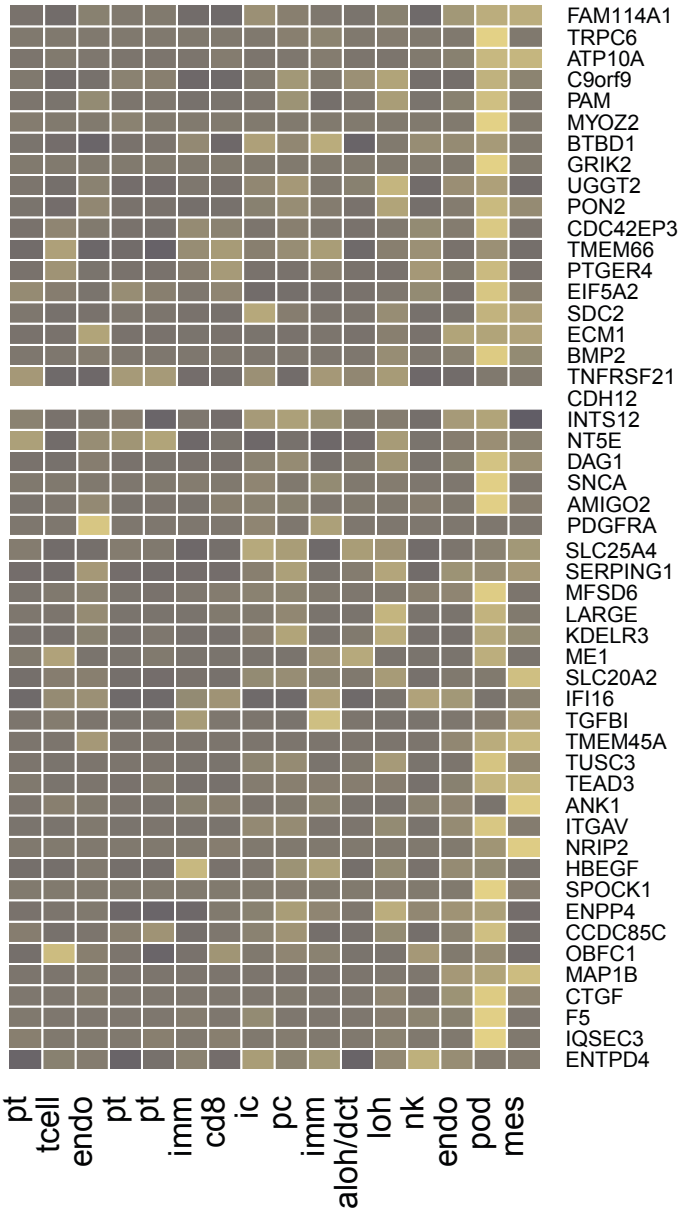
Aspect	GoTerm name	GoTerm id	# Query Genes in ...	Total # of Genes i...	Query Genes in Te...	p-value
biological_process	circulatory syste...	GO:0072359	14	1037	ADM, DAG1, PDL...	9.456E-3
biological_process	extracellular mat...	GO:0030198	9	393	APLP1, DAG1, S...	1.106E-2
biological_process	extracellular stru...	GO:0043062	9	394	APLP1, DAG1, S...	1.129E-2
biological_process	developmental p...	GO:0032502	37	6338	ADM, APLP1, DA...	2.242E-2
biological_process	negative regulati...	GO:0060192	3	16	ANGPTL4, PLA2...	4.919E-2
cellular_component	collagen-containi...	GO:0062023	10	423	APLP1, DAG1, S...	2.256E-3
cellular_component	extracellular mat...	GO:0031012	11	562	APLP1, DAG1, S...	3.811E-3
cellular_component	main axon	GO:0044304	5	62	DAG1, MYO1D, ...	4.494E-3
cellular_component	extracellular region	GO:0005576	31	4399	ADM, APLP1, DA...	4.778E-3
cellular_component	integrin alphav-...	GO:0034686	2	2	ITGAV, ITGB8,	2.971E-2
cellular_component	extracellular space	GO:0005615	25	3374	ADM, DAG1, IQG...	4.006E-2
molecular_function	double-stranded...	GO:0036121	2	2	CHD3, ANXA1,	2.971E-2

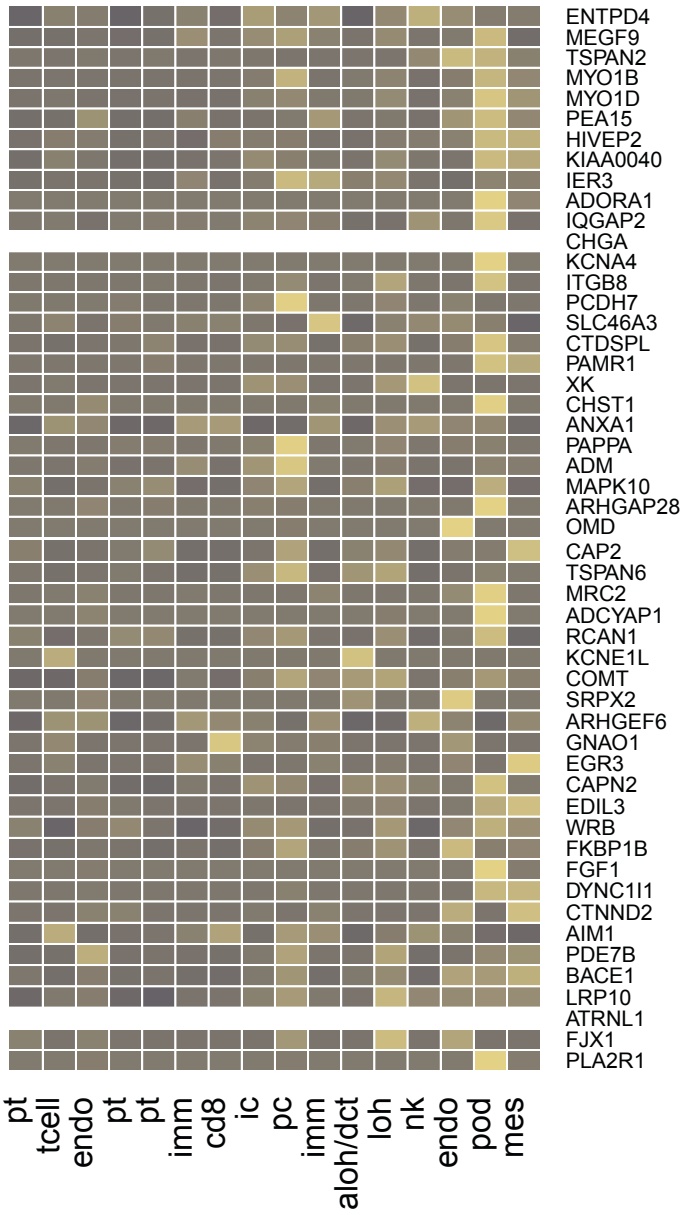
B.

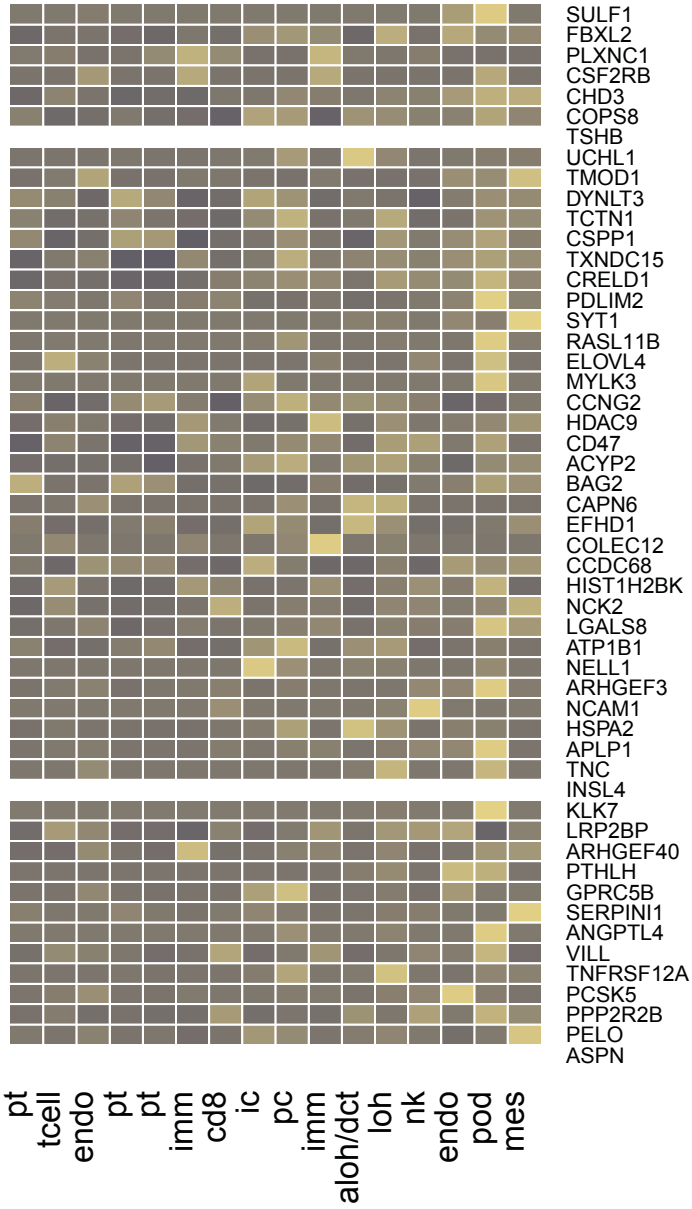


Module	Top Term	Qval	Genes
M1	regulation of small molecule metabolic process	0.003	8
M2	actin cytoskeleton organization	0.003	3
M3	regulation of response to external stimulus	0.006	9

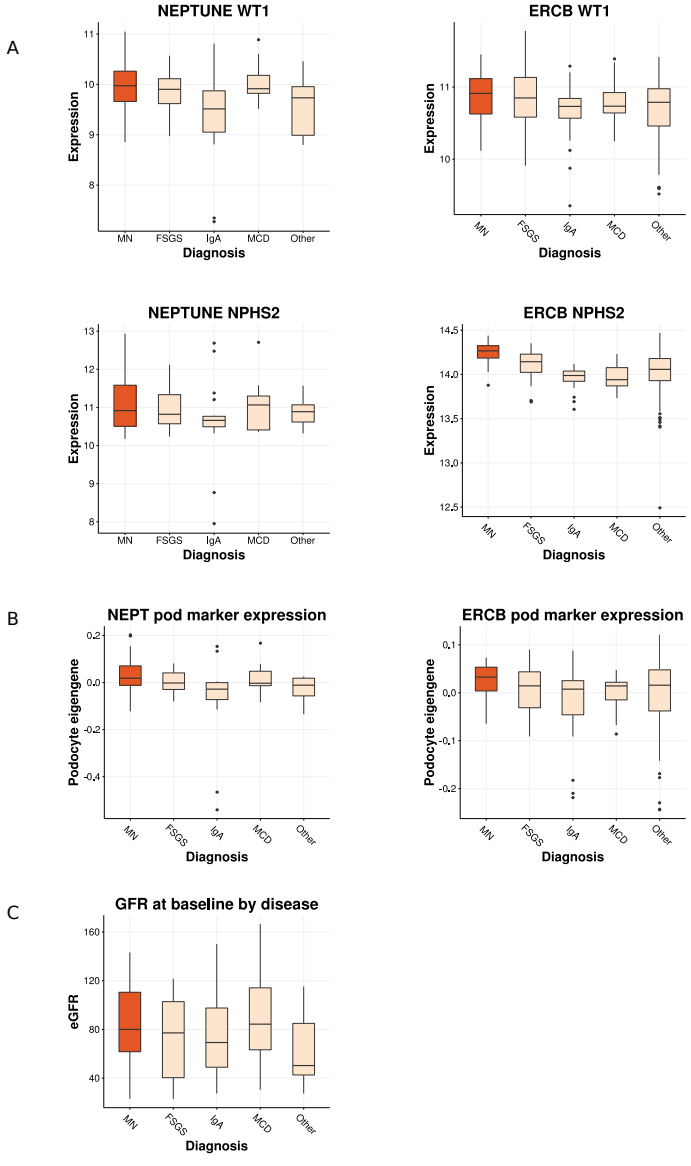
Supplemental Figure 10. Expression of MN-specific genes across single-cell clusters. Cell-type specific expression of MN-specific genes expression was examined across single-cell clusters. For each gene, mean expression in each cluster was identified and scaled across cell types. Many MN-specific genes are podocyte expressed. aloh/dct, ascending loop of Henle/distal convoluted tubule; cd8, CD8+ T cell; endo, endothelial cell; ic=intercalated cell; imm, immune cell; loh, loop of Henle; mes, mesangial/vascular smooth muscle cell; pc, principal cell; pod, podocyte; pt, proximal tubule; tcell, T cell.


FAM114A1
TRPC6
ATP10A
C9orf9
PAM
MYOZ2
BTBD1
GRIK2
UGGT2
PON2
CDC42EP3
TMEM66
PTGER4
EIF5A2
SDC2
ECM1
BMP2
TNFRSF21
CDH12
INTS12
NT5E
DAG1
SNCA
AMIGO2
PDGFRA
SLC25A4
SERPING1
MFSD6
LARGE
KDELR3
ME1
SLC20A2
IFI16
TGFBI
TMEM45A
TUSC3
TEAD3
ANK1
ITGAV
NRIP2
HBEGF
SPOCK1
ENPP4
CCDC85C
OBFC1
MAP1B
CTGF
F5
IQSEC3
ENTPD4
pt
tcell
endo
pt
pt
imm
cd8
ic
pc
imm
aloh/dct
loh
nk
endo
pod
mes

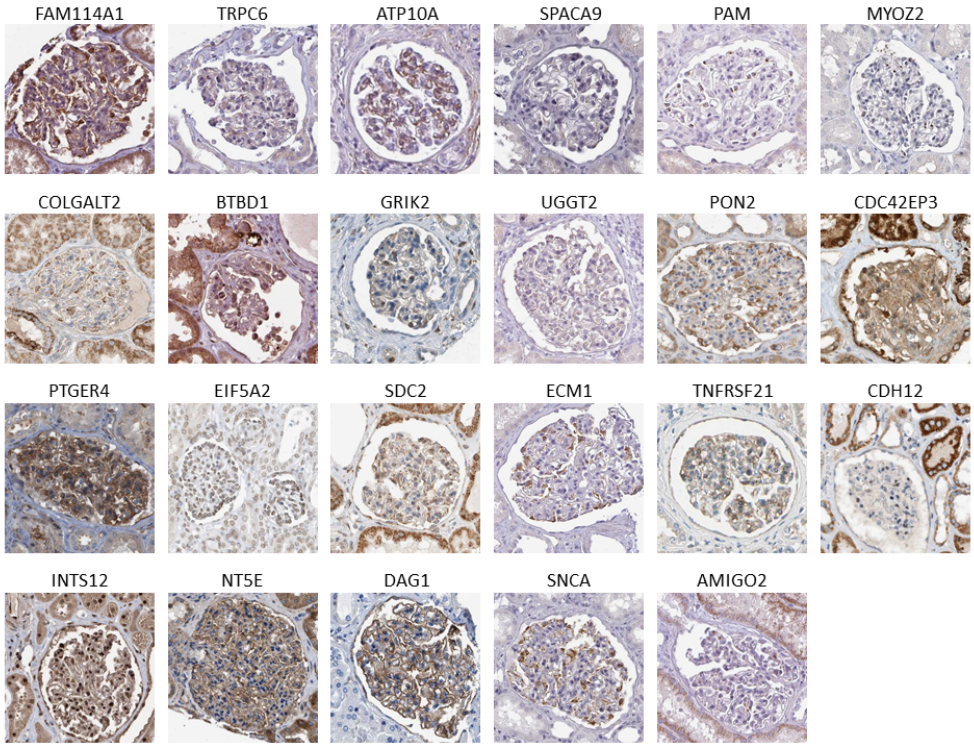

ENTPD4
MEGF9
TSPAN2
MYO1B
MYO1D
PEA15
HIVEP2
KIAA0040
IER3
ADORA1
IQGAP2
CHGA
KCNA4
ITGB8
PCDH7
SLC46A3
CTDSPL
PAMR1
XK
CHST1
ANXA1
PAPPA
ADM
MAPK10
ARHGAP28
OMD
CAP2
TSPAN6
MRC2
ADCYAP1
RCAN1
KCNE1L
COMT
SRPX2
ARHGEF6
GNAO1
EGR3
CAPN2
EDIL3
WRB
FKBP1B
FGF1
DYNC1I1
CTNND2
AIM1
PDE7B
BACE1
LRP10
ATRNL1
FJX1
PLA2R1
pt
tcell
endo
pt
pt
imm
cd8
ic
pc
imm
aloh/dct
loh
nk
endo
pod
mes


SULF1
FBXL2
PLXNC1
CSF2RB
CHD3
COPS8
TSHB
UCHL1
TMOD1
DYNLT3
TCTN1
CSPP1
TXNDC15
CRELD1
PDLIM2
SYT1
RASL11B
ELOVL4
MYLK3
CCNG2
HDAC9
CD47
ACYP2
BAG2
CAPN6
EFHD1
COLEC12
CCDC68
HIST1H2BK
NCK2
LGALS8
ATP1B1
NELL1
ARHGEF3
NCAM1
HSPA2
APLP1
TNC
INSL4
KLK7
LRP2BP
ARHGEF40
PTHLH
GPRC5B
SERPINI1
ANGPTL4
VILL
TNFRSF12A
PCSK5
PPP2R2B
PELO
ASPN
pt
tcell
endo
pt
pt
imm
cd8
ic
pc
imm
aloh/dct
loh
nk
endo
pod
mes

Supplemental Figure 11. Expression of canonical podocyte genes WT1 and NPHS2, podocyte eigengene, and eGFR across disease types and cohorts. (A) Comparison of expression of key podocyte markers WT1 (upper panels) and NPHS2 (lower panels) across diseases, in NEPTUNE (left) and ERCB (right) cohorts. (B) Podocyte eigengene across diseases in NEPTUNE and ERCB. Higher eigengene values suggest increased podocyte proportion. The podocyte eigengene was identified using the CellCODE framework, using as marker genes the top 20 markers associated with cell type clusters in the human tumor nephrectomy single-cell data from Gillies *et al.* (2018). (C) eGFR in MN subjects in NEPTUNE is not significantly higher than in other glomerulonephropathies (p=0.09 for MN *vs.* all others by t-test; p=0.69 for MN vs. MCD). Red boxes, MN; beige boxes, diagnoses other than MN.



Supplemental Figure 12. Human Protein Atlas images showing presence or absence of glomerular expression for 23 of the 25 top MN classifier genes. Representative images from the Human Protein Atlas database (www.proteinatlas.org) are shown. See Supplemental Methods for additional details.



C. Supplemental Tables (please refer to separate Excel spreadsheet)

Supplemental Table 1. Anonymized IDs for original NEPTUNE, ERCB, expanded NEPTUNE, validation, and high-proteinuria participants included in study, and PLA2R statuses for NEPTUNE participants.

Supplemental Table 2. Baseline clinical characteristics of NEPTUNE, ERCB, and validation cohort participants included in the study, tabulated by diagnosis as well as by MN vs. all other diagnoses. Baseline characteristics of NEPTUNE participants included in study, including age, sex, race/ethnicity, clinical characteristics, are reported, as well as clinical outcomes. For numeric characteristics, the median and interquartile range are presented. For categorical characteristics, the number and percentage are presented. The clinical metadata available for the ERCB subjects is more limited.

Supplemental Table 3. Mean AUCs for disease-specific classifiers. The disease-specific random forest classifiers use as features all glomerular expression data from NEPTUNE; all glomerular expression data from ERCB; all tubulointerstitial expression data from NEPTUNE; all tubulointerstitial expression data from ERCB; WGCNA eigengenes of glomerular NEPTUNE expression data; and WGCNA eigengenes of tubulointerstitial expression data, respectively. RF, random forest; AUC, area under the receiver operating characteristic curve.

Supplemental Table 4. MN-specific gene list and overlap with genes previously reported to be differentially expressed in a Heymann nephritis rat model. List of genes that are differentially expressed in MN compared to other diseases in both the ERCB and NEPTUNE cohorts, ranked by Gini importance in the random forest classifier. The rightmost column lists the genes overlapping with those found by Hauser et al. (2009) in a rat experimental model of MN.

Supplemental Table 5. Module enrichments within the MN-specific gene list. MN-specific genes were grouped into modules based on their connectivity in a kidney-specific functional network. Functional enrichments were identified for each module.

Supplemental Table 6. MN-specific genes overlapping with podocyte markers from single-cell and single-nucleus RNA sequencing studies. Multiple podocyte-specific lists are considered for overlap with the MN-specific gene list: 1) Park *et al.*, 2018 (mouse); 2) Gillies *et al*., 2018 (human); 3) Lake *et al*., 2019 (human); 4) Menon *et al*., 2020 (human); 5) Stewart *et al*., 2019 (human). See main article for full references.

Supplemental Table 7. Module enrichments within the podocyte-expressed MN-specific gene list. Podocyte-expressed, MN-specific genes (Supplemental Table 6) were grouped into modules based on their connectivity in a kidney-specific functional network. Functional enrichments were identified for each module.

Supplemental Table 8. Comparison of expression of canonical podocyte marker genes, eigengenes, and GFRs for participants with MN vs. all other diagnoses and participants with MN vs. participants with MCD. We compared the expression of canonical podocyte marker genes (NPHS2, PODXL, WT1, and COL4A3), the podocyte eigengene, and GFR between participants with MN and all other diseases, as well as between participants with MN and participants with MCD in the NEPTUNE and ERCB cohorts. Each p-value was computed using a one-sided t-test (testing whether the MN value was significantly greater than the other value). GFR, glomerular filtration rate; SPV, podocyte surrogate proportion variable.

Supplemental Table 9. Metadata for the Human Protein Atlas images used to generate Supplemental Figure 12. The age and sex associated with the kidney tissue, the antibody used for staining, and the URL of the image are provided for each image selected for the composite.