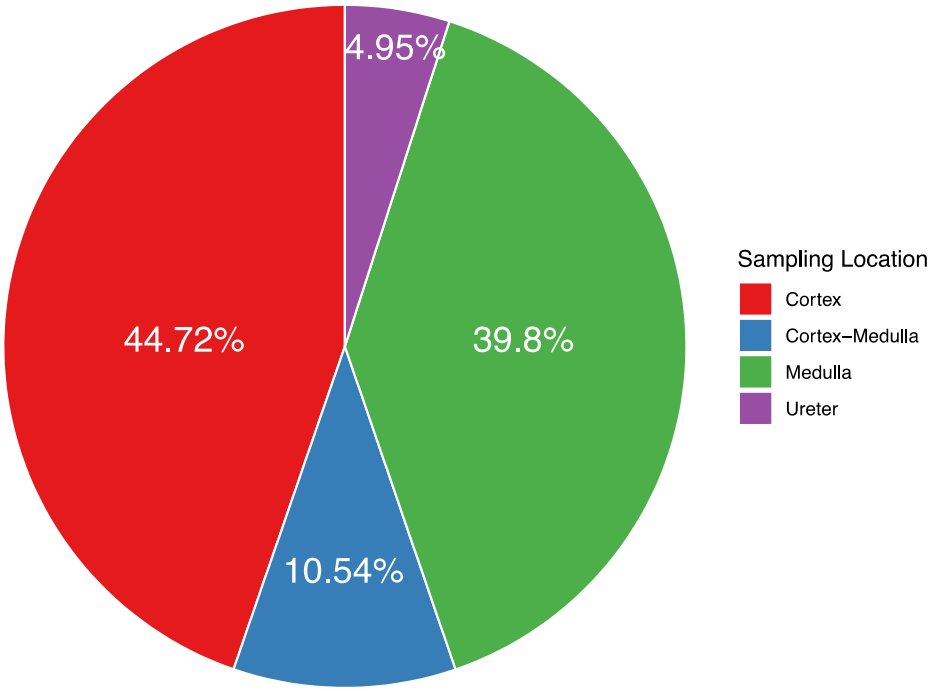
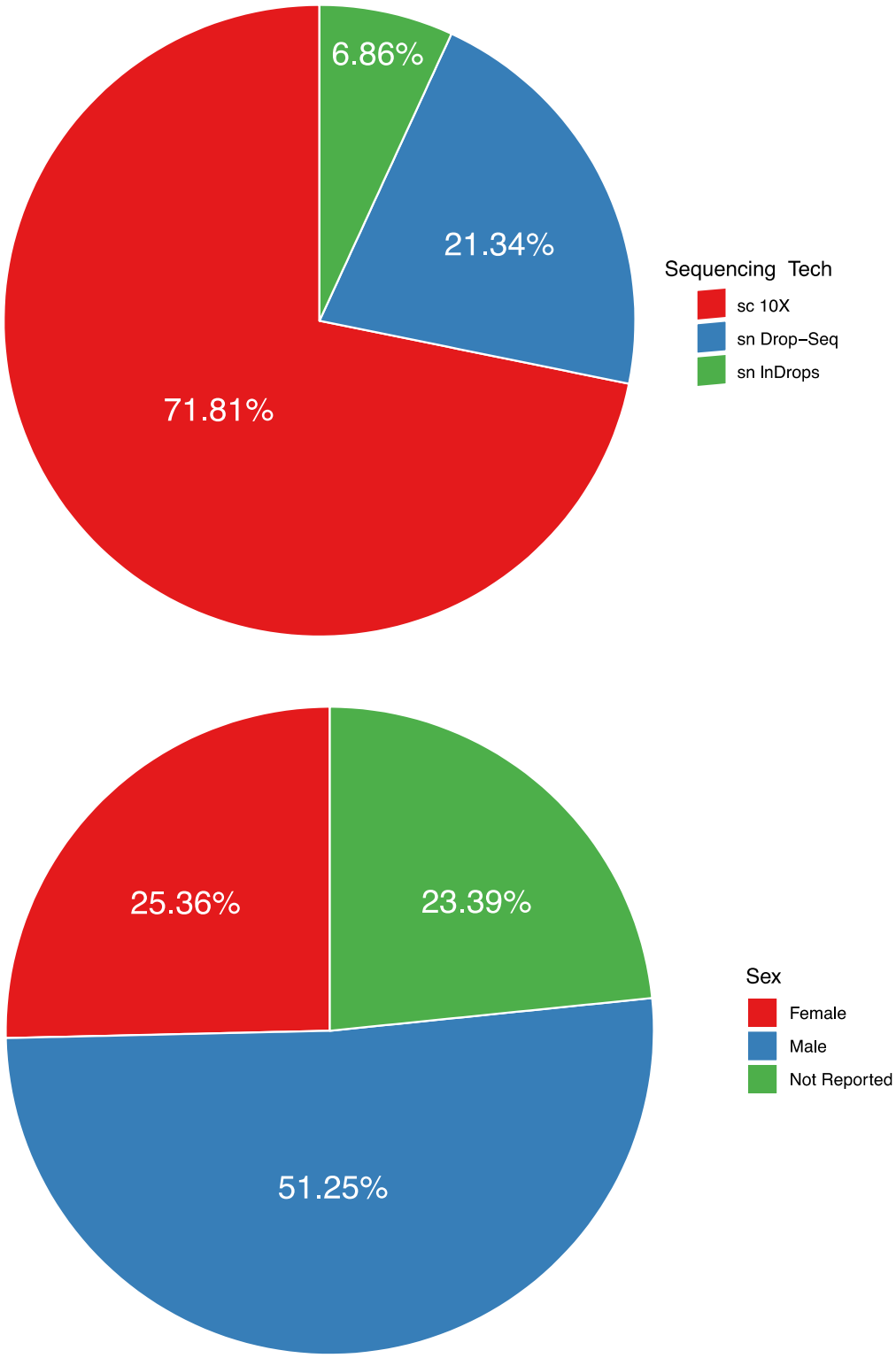
Supplementary Figures

Figure S1. Distribution of renal cells in our combined cohort by (a) known sampling location, (b) sequencing technology, and (c) donor sex. Cells with unknown sampling location are excluded from this figure.

(A)



(B)

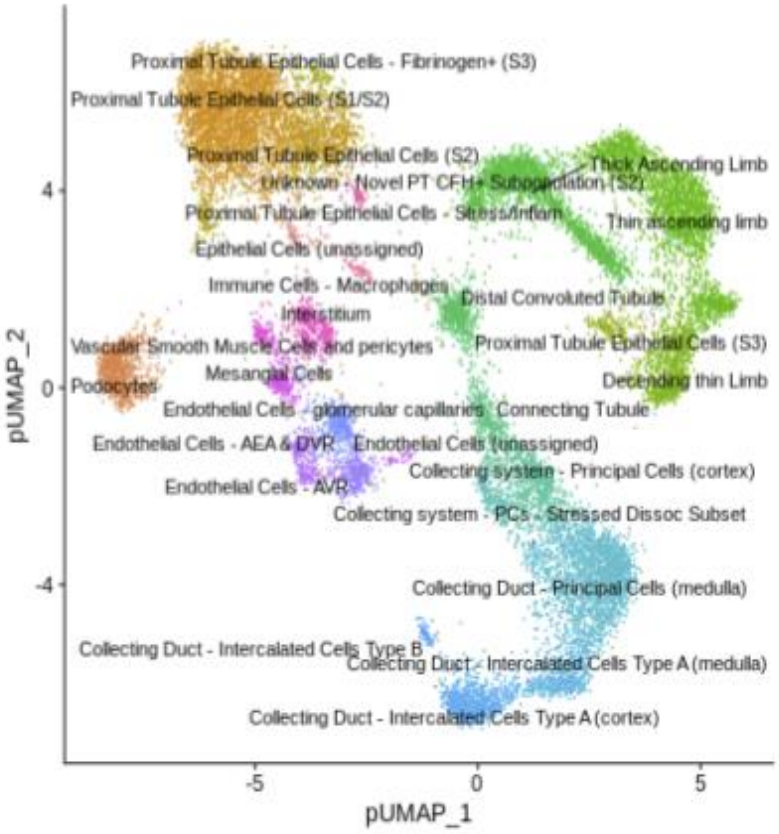

6.86%
21.34%
71.81%
Sequencing Tech
sc 10X
sn Drop−Seq
sn InDrops

(C)

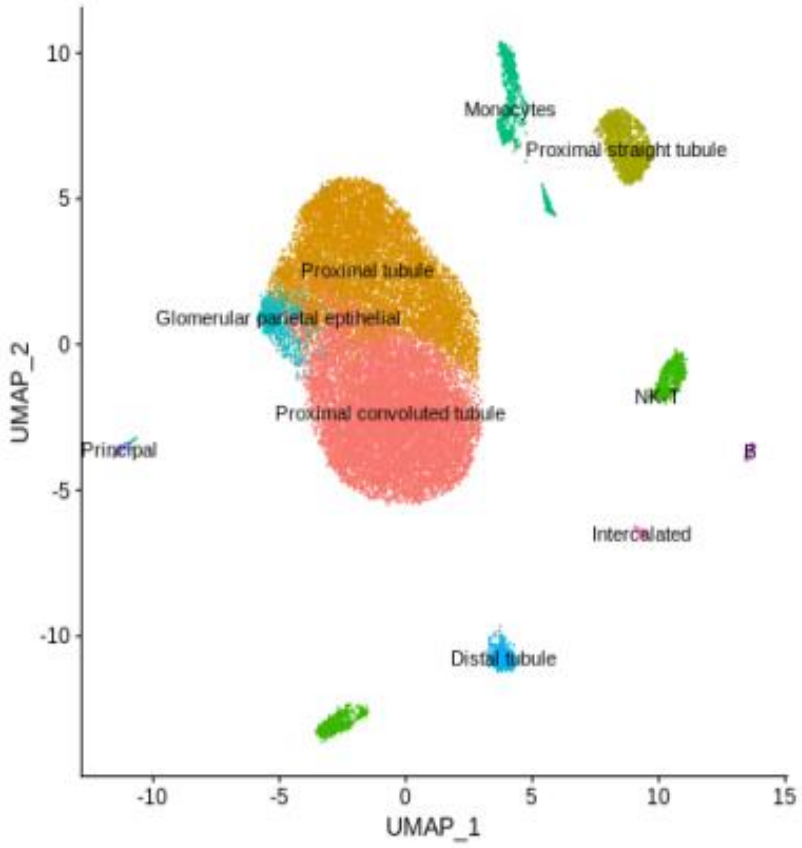
25.36%
23.39%
51.25%
Sex
Female
Male
Not Reported

Figure S2: Uniform manifold approximation and projection for dimension reduction (UMAP) and t-distributed stochastic neighbor embedding (t-SNE) visualizations of each dataset generated using the code provided by authors in their publication or sent to us directly for (A) Lake et al., (B) Liao et al., (C) Menon et al., (D) Wu et al., (E) Young et al.

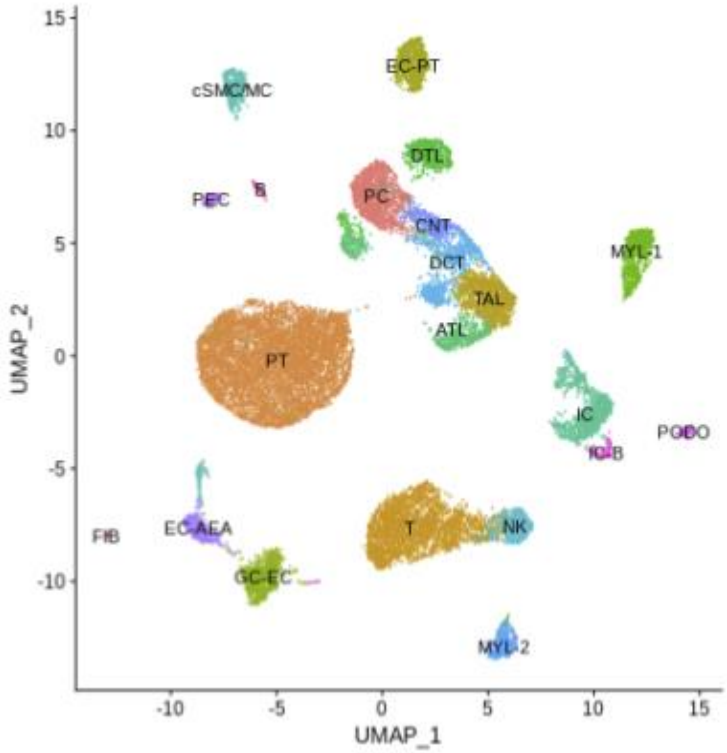
(A)



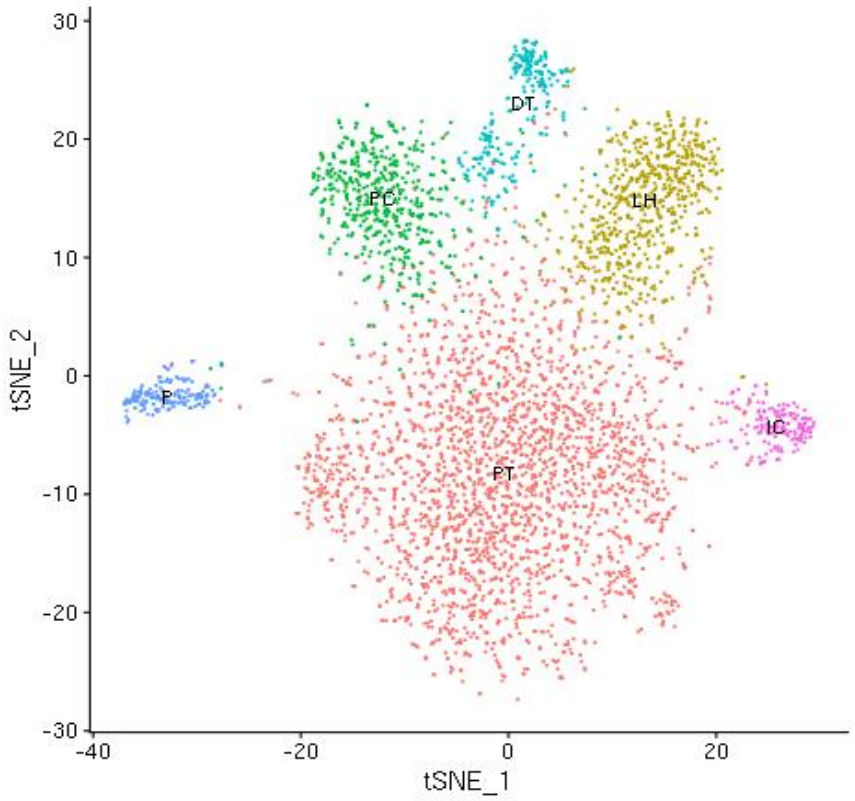
(B)


Monocytes
Proximal straight tubule
Proximal tubule
Glomerular parietal epithelial
Proximal convoluted tubule
NK-T
Principal
B
Intercalated
Distal tubule
UMAP_2
UMAP_1
10
5
0
-5
-10
-10
-5
0
5
10
15

(C)


EC-PT
cSMC/MC
DTL
PEC
B
PC
CNT
DCT
TAL
ATL
MYL-1
PT
IC
POD
IC-B
FIB
EC-AEA
T
NK
GC-EC
MYL-2
UMAP_2
UMAP_1
15
10
5
0
-5
-10
-10
-5
0
5
10
15

(D)


DT
PC
LH
P
IC
PT
tSNE_2
tSNE_1
30
20
10
0
-10
-20
-30
-40
-20
0
20

(E)

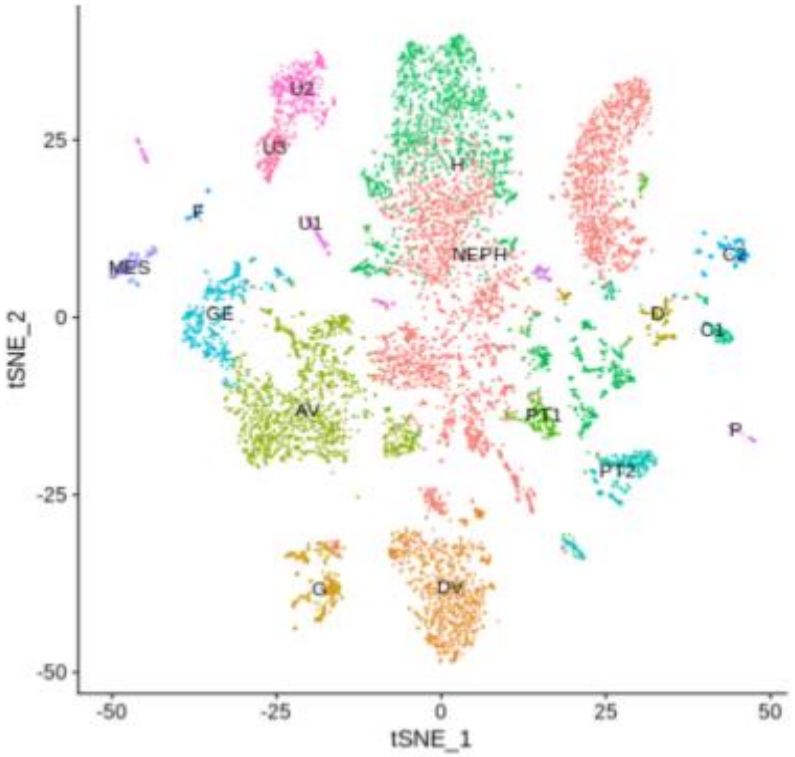
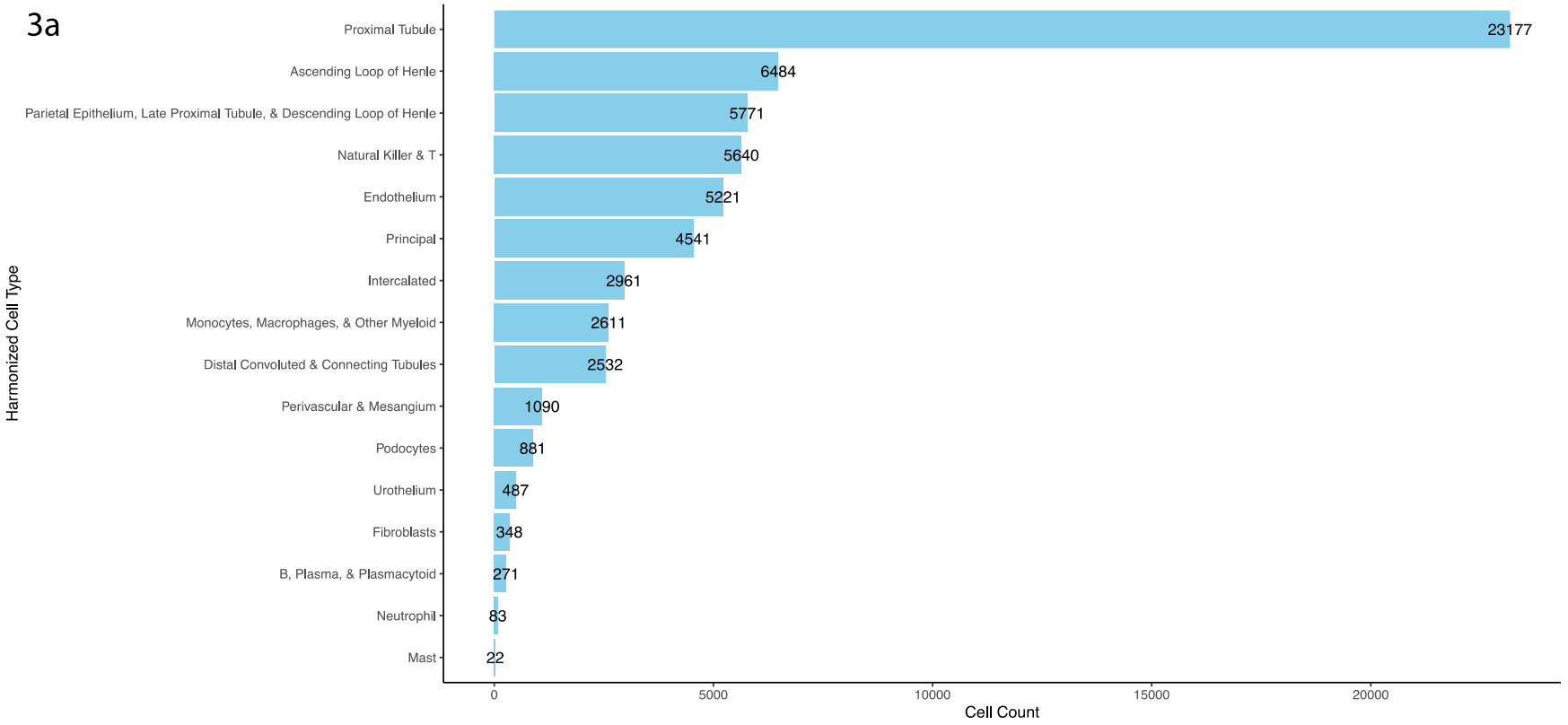

U2
U3
H
F
U1
MES
NEPH
GE
D
C1
AV
PT1
P
PT2
C
DV
tSNE_2
tSNE_1
25
0
-25
-50
-50
-25
0
25
50

Figure S3. Barcharts showing (a) the distribution of cell counts by harmonized cell type and (b) the proportion of cells from each study for each harmonized cell type.

(A)



(B)

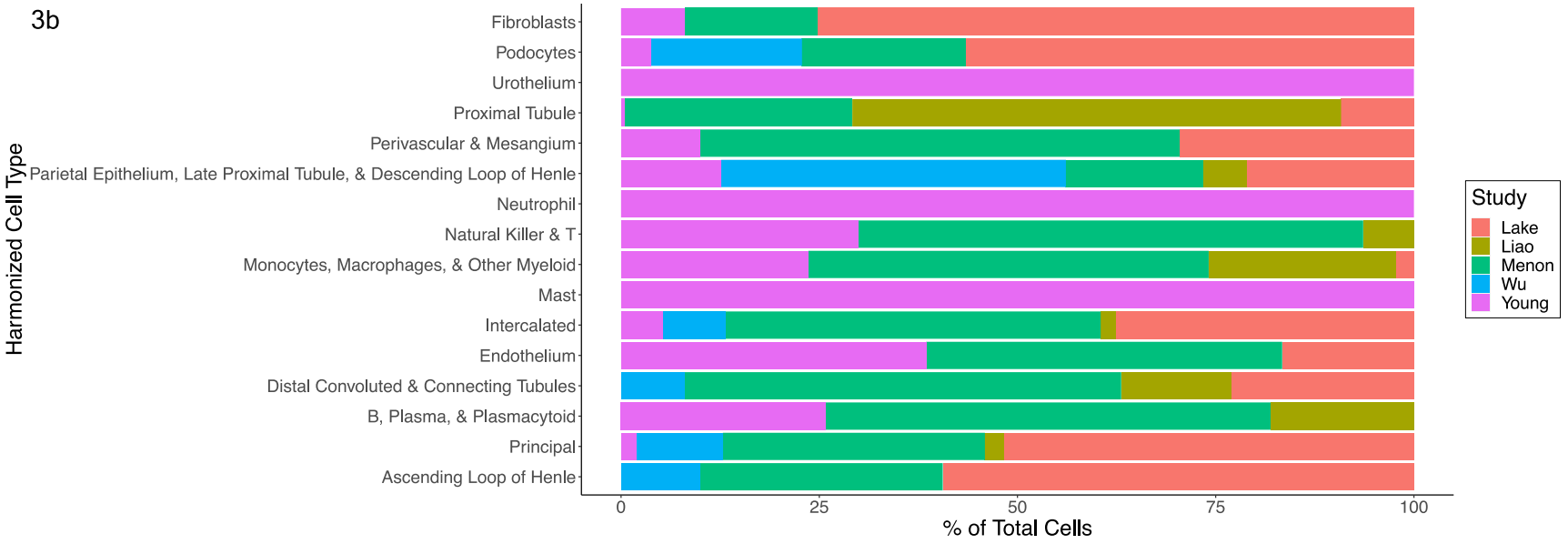
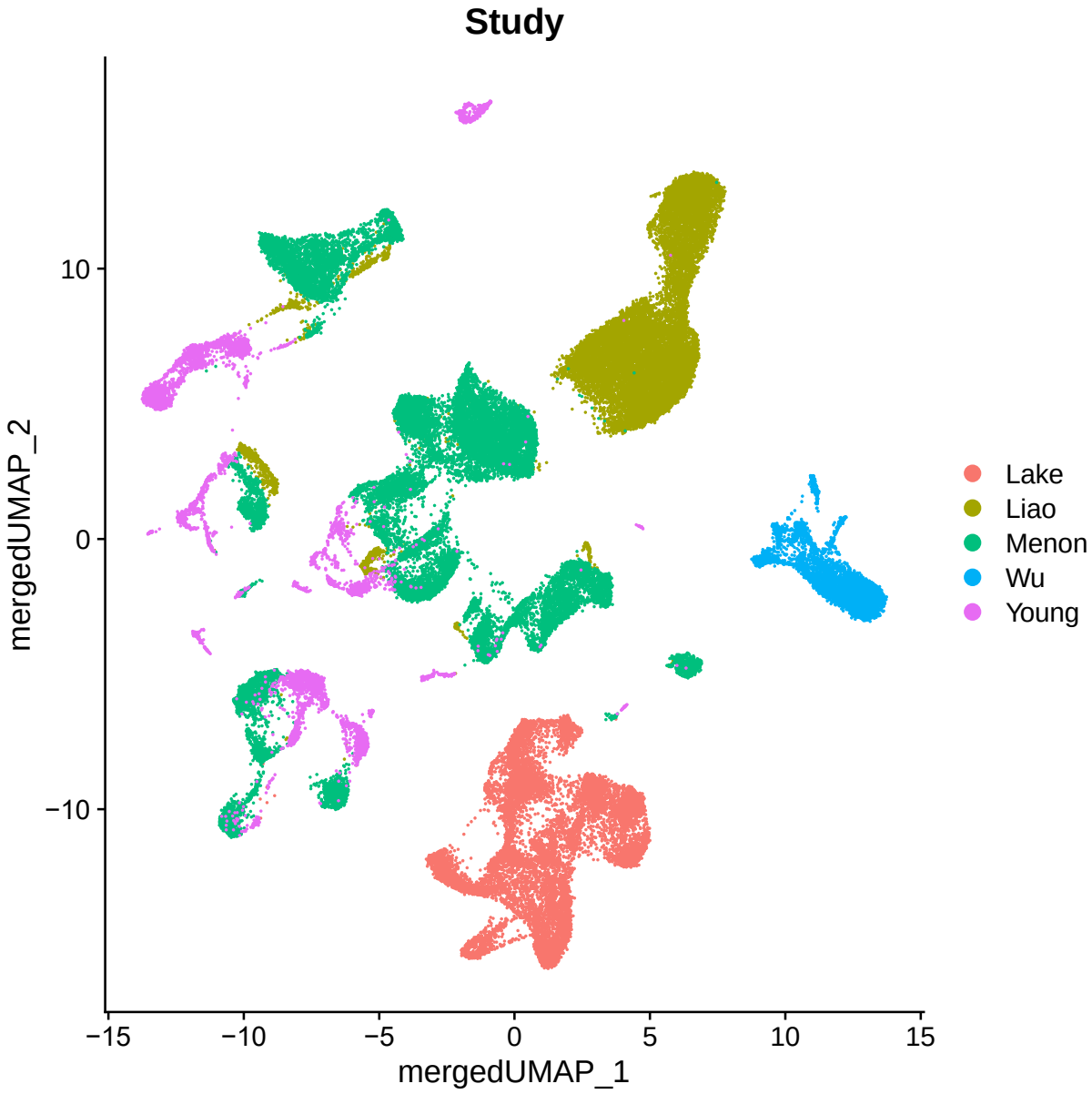
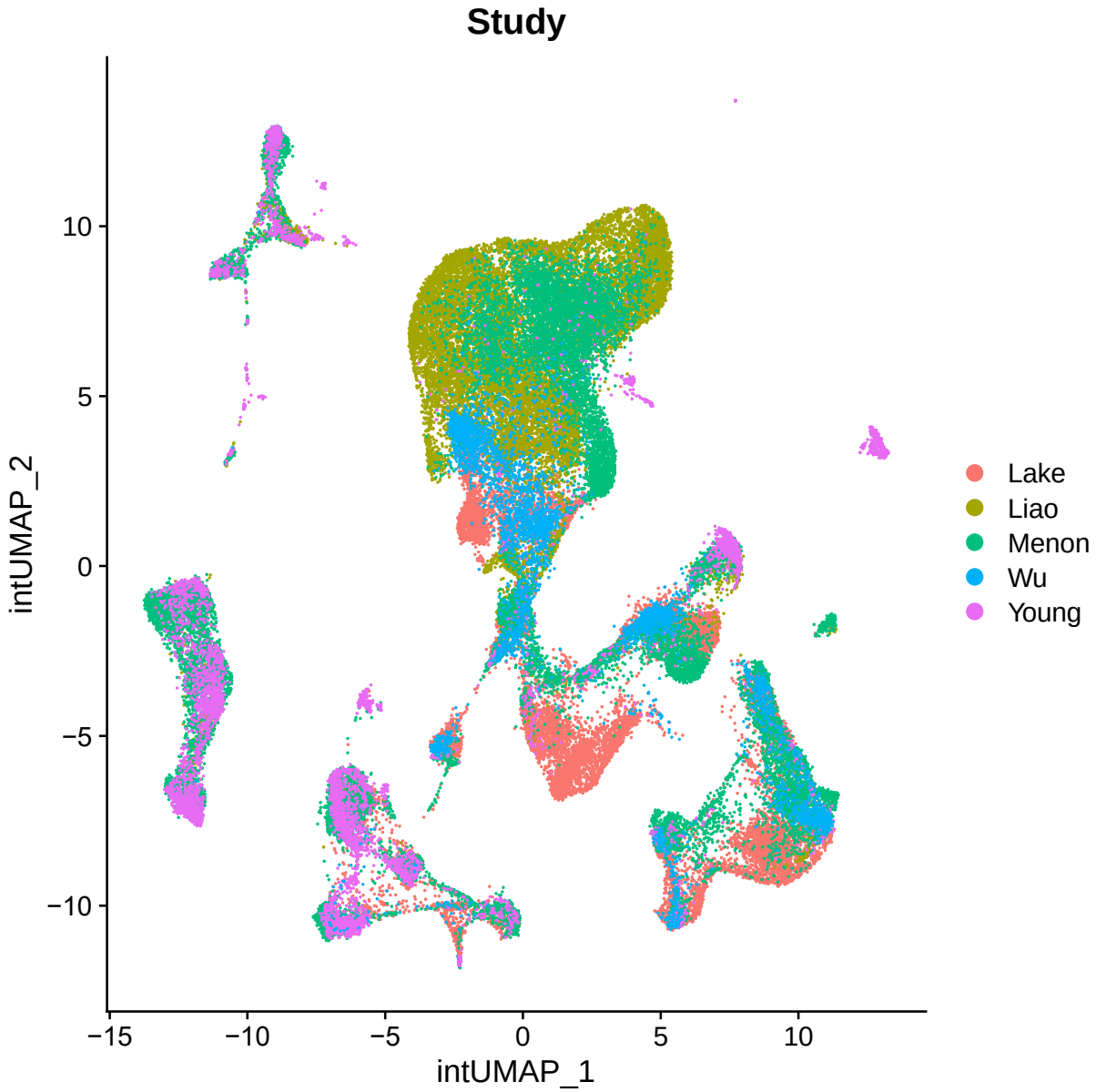

3b
Harmonized Cell Type
Fibroblasts
Podocytes
Urothelium
Proximal Tubule
Perivascular & Mesangium
Parietal Epithelium, Late Proximal Tubule, & Descending Loop of Henle
Neutrophil
Natural Killer & T
Monocytes, Macrophages, & Other Myeloid
Mast
Intercalated
Endothelium
Distal Convoluted & Connecting Tubules
B, Plasma, & Plasmacytoid
Principal
Ascending Loop of Henle
0
25
50
75
100
% of Total Cells
Study
Lake
Liao
Menon
Wu
Young

Figure S4. UMAP visualizations of the combined dataset (a) before batch correction, (b) after batch correction, and (c) after batch correction colored by harmonized cell type.

(A)



(B)


Study
intUMAP_2
intUMAP_1
10
5
0
−5
−10
−15
−10
−5
0
5
10
Lake
Liao
Menon
Wu
Young

(C)

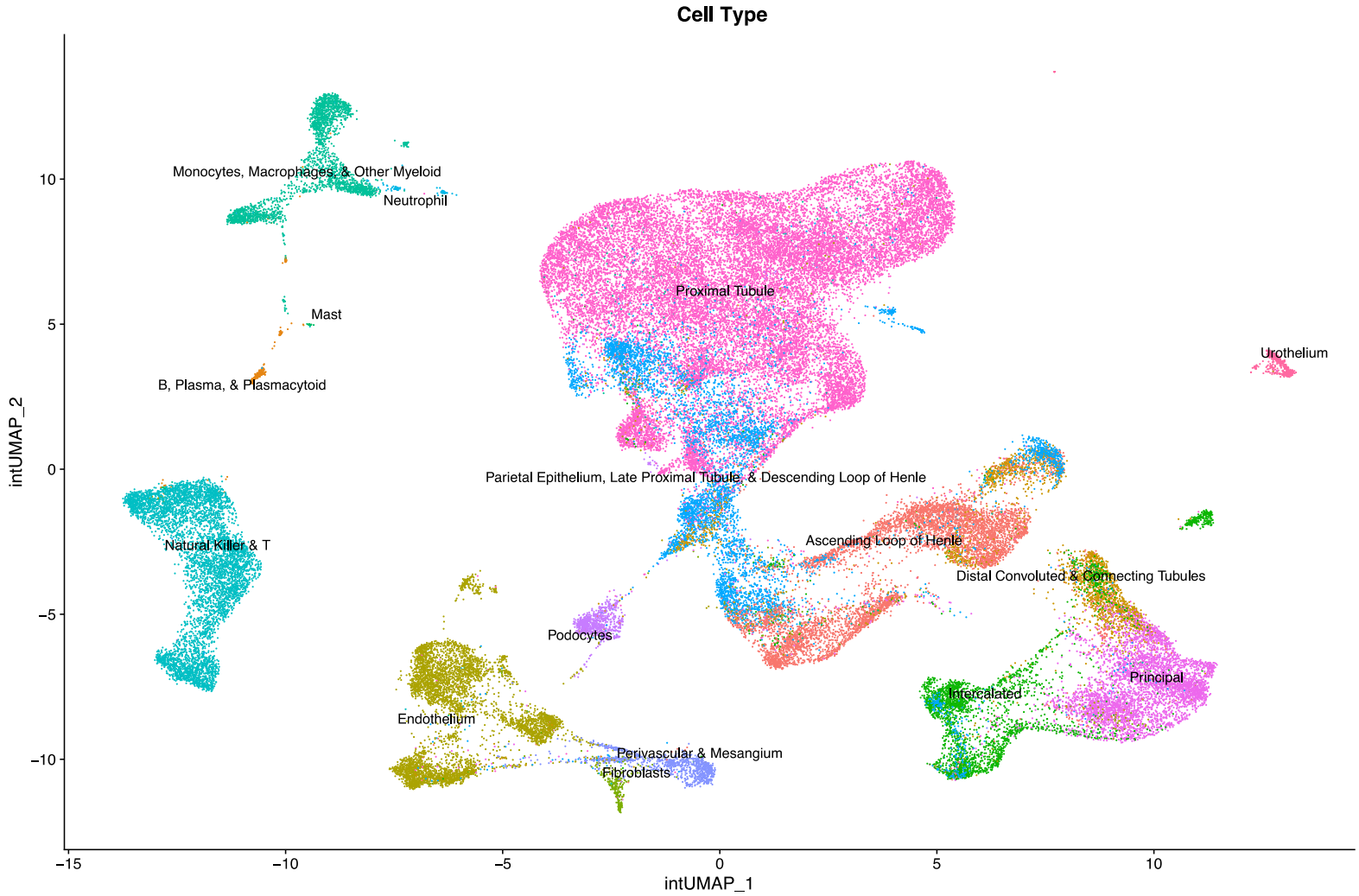

Cell Type
Monocytes, Macrophages, & Other Myeloid
Neutrophil
Mast
B, Plasma, & Plasmacytoid
Proximal Tubule
Urothelium
Parietal Epithelium, Late Proximal Tubule, & Descending Loop of Henle
Natural Killer & T
Ascending Loop of Henle
Distal Convoluted & Connecting Tubules
Podocytes
Principal
Intercalated
Endothelium
Perivascular & Mesangium
Fibroblasts
intUMAP_2
intUMAP_1
10
5
0
−5
−10
−15
−10
−5
0
5
10

Figure S5: Schematic of SVM outlier detection with 4,257 cells removed from further analysis due to poor alignment with their original annotated cell type. UMAP of cells within the “Proximal Tubule” harmonized cell type before (left) and after (right) SVM outlier detection. Cells circled in red are poorly aligned with the remaining cells of their harmonized cell type, a trend observed with all other harmonized cell types.

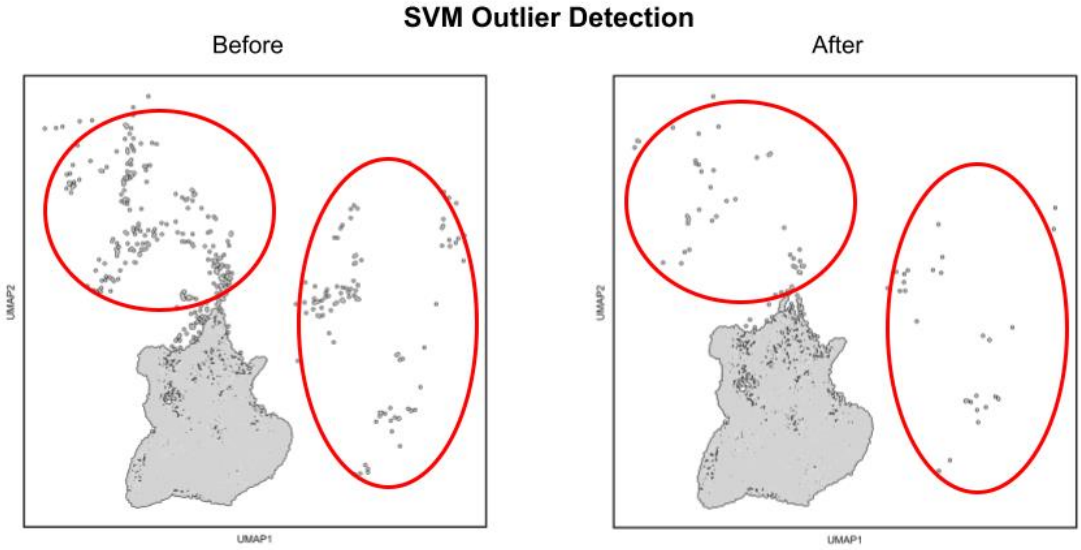
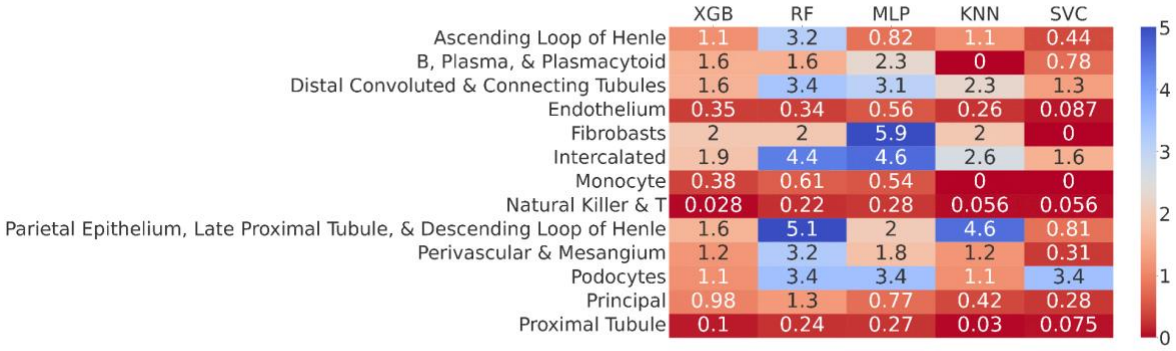


Table S1. Distribution of 57,864 renal cells used in our analyses, following SVM-based exclusion of low-quality cells.

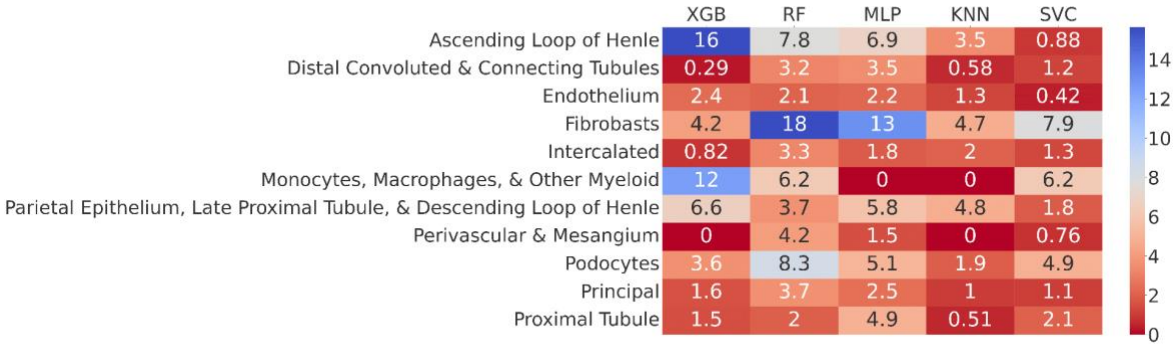
Contribution to Harmonized Cell Type Counts by Study						
Harmonized Cell Type	**Lake et al.**	**Liao et al.**	**Menon et al.**	**Wu et al.**	**Young et al.**	**Total**
Ascending Loop of Henle	3620	0	1835	581	0	6036
B, Plasma, & Plasmacytoid	0	42	129	0	69	240
Distal Convoluted & Connecting Tubules	344	309	745	183	0	1581
Endothelium	717	0	2308	0	1956	4981
Fibroblasts	191	0	51	0	28	270
Intercalated	976	48	1336	215	153	2728
Mast	0	0	0	0	22	22
Monocytes, Macrophages, & Other Myeloid	16	490	1307	0	616	2429
Natural Killer & T	0	351	3582	0	1672	5605
Neutrophil	0	0	0	0	77	77
Parietal Epithelium, Late Proximal Tubule, & Descending Loop of Henle	980	71	738	2230	568	4587
Perivascular & Mesangium	265	0	654	0	107	1026
Podocytes	468	0	174	138	33	813
Principal	2157	82	1429	426	59	4153
Proximal Tubule	1950	14157	6639	0	84	22830
Urothelium	0	0	0	0	486	486

Figure S6. Heatmap of each classifier's rejection rate on (a) Menon et al., (b) Lake et al., (c) Liao et al. and (d) Wu et al with respect to each cell type.

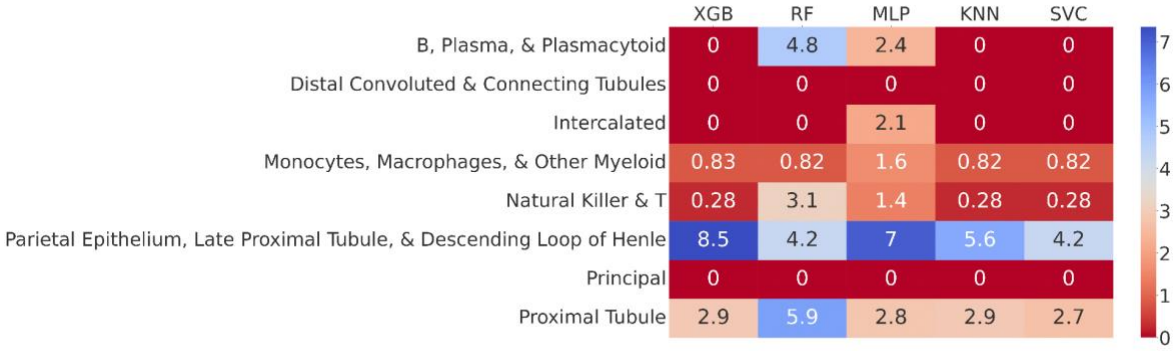
(a)



(b)



(c)



(d)

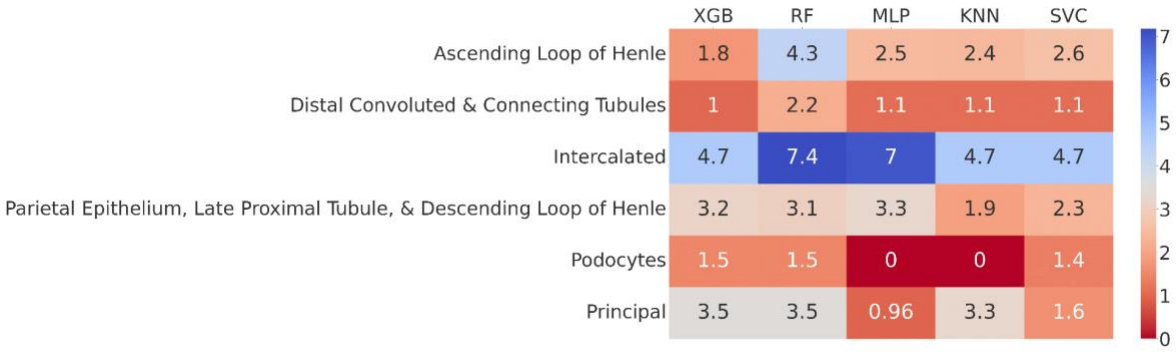

XGB
RF
MLP
KNN
SVC
Ascending Loop of Henle
1.8
4.3
2.5
2.4
2.6
Distal Convoluted & Connecting Tubules
1
2.2
1.1
1.1
1.1
Intercalated
4.7
7.4
7
4.7
4.7
Parietal Epithelium, Late Proximal Tubule, & Descending Loop of Henle
3.2
3.1
3.3
1.9
2.3
Podocytes
1.5
1.5
0
0
1.4
Principal
3.5
3.5
0.96
3.3
1.6
7
6
5
4
3
2
1
0

Figure S7. Heatmap of each classifier's F1 score on Liao et al. with respect to each cell type.

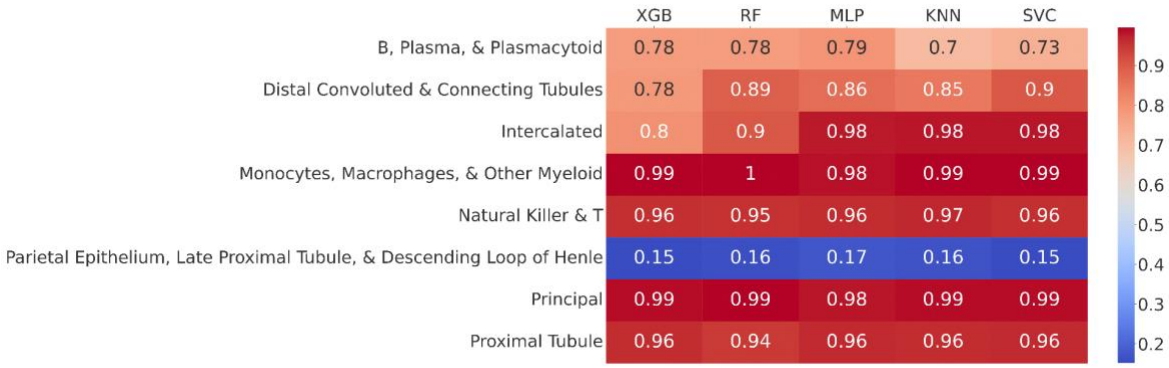


Figure S8: Schematic of our study snakemake pipeline.

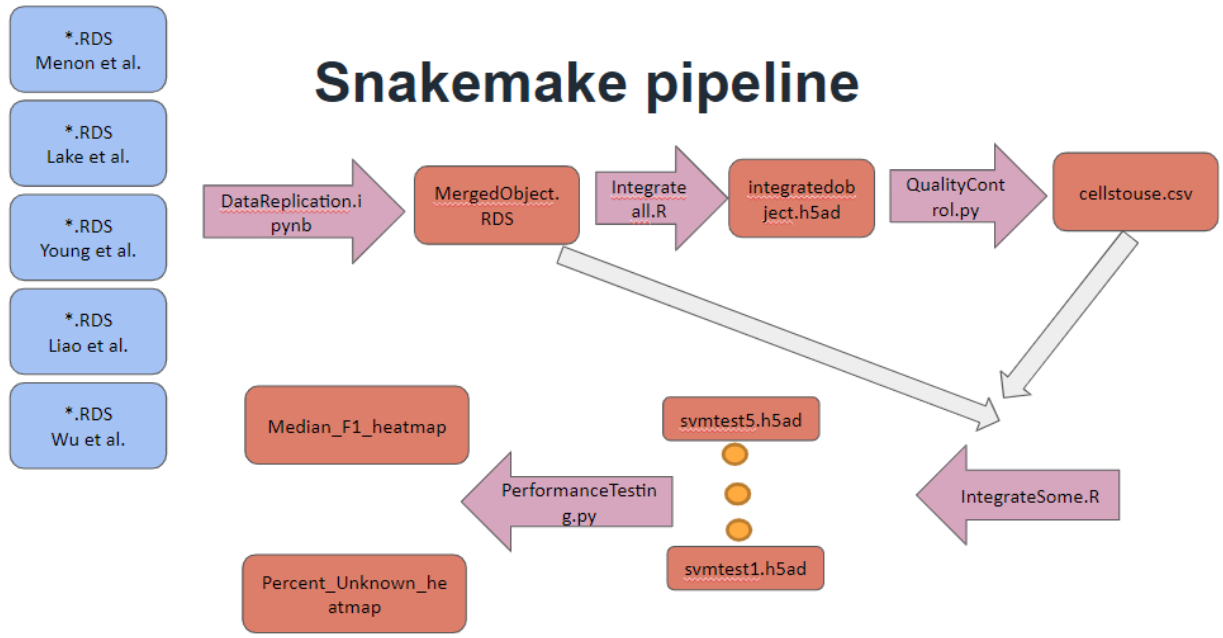
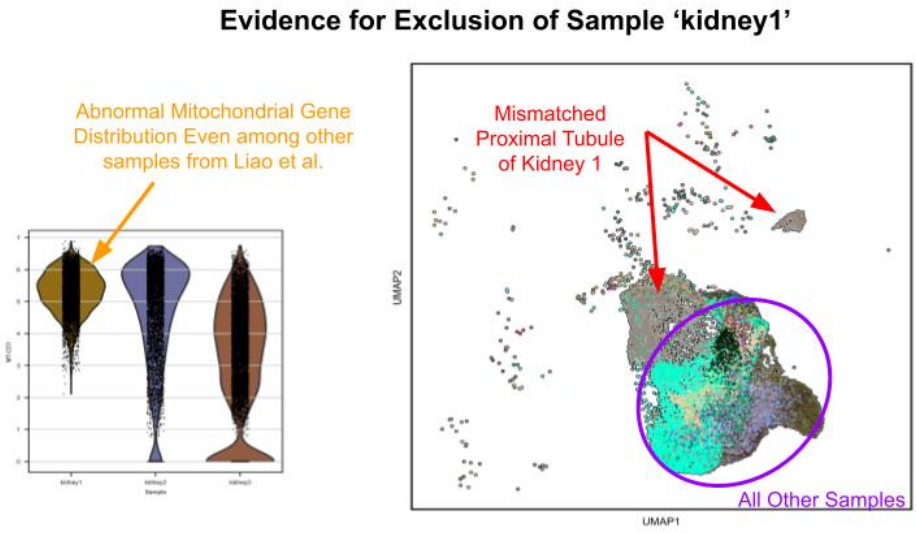
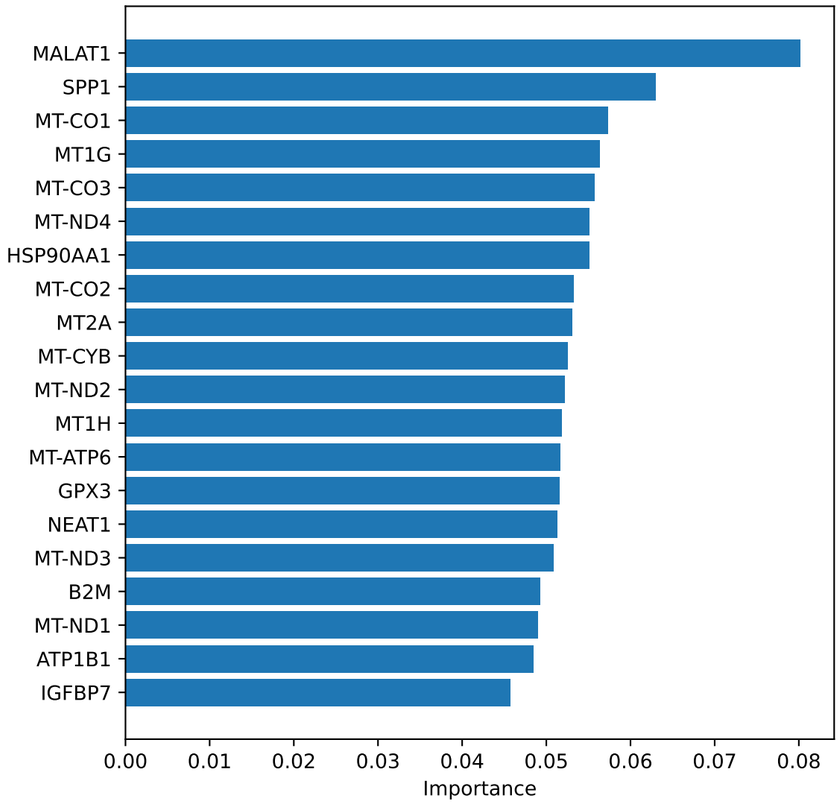
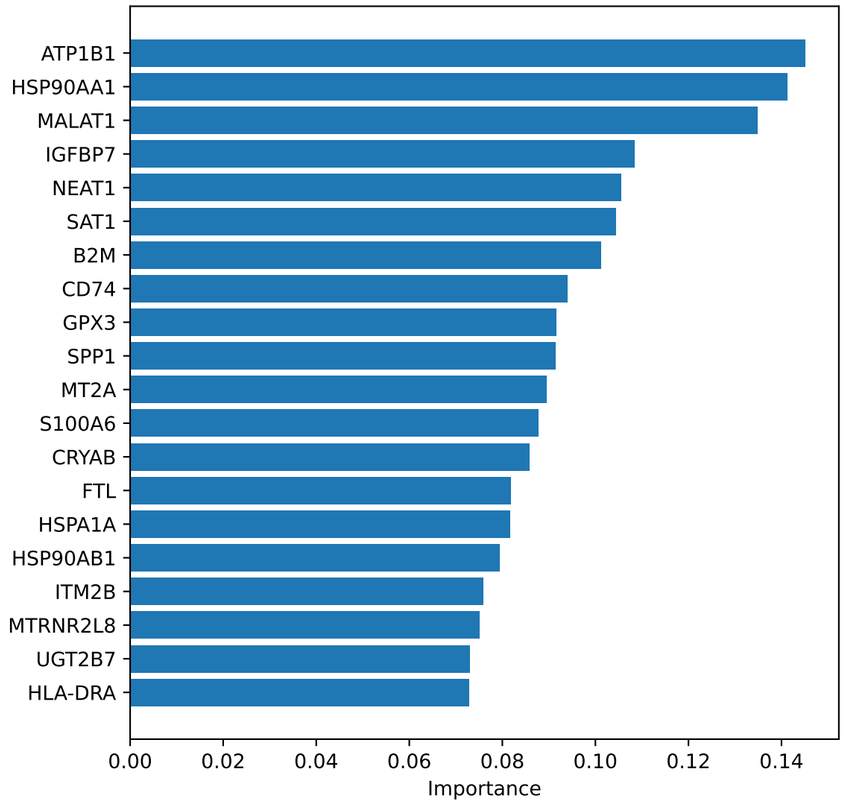
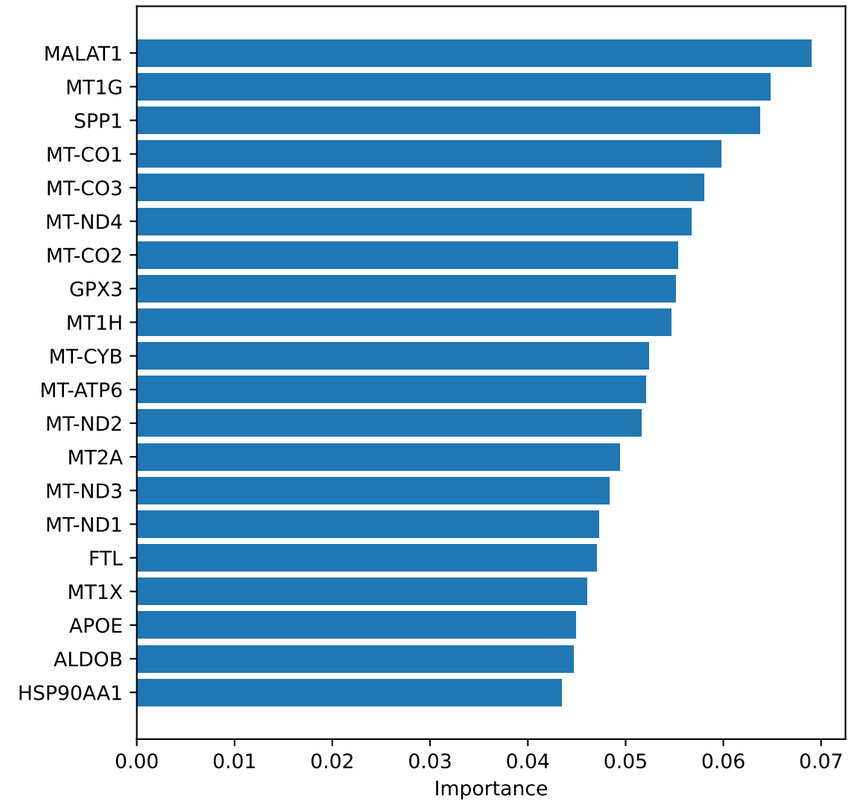


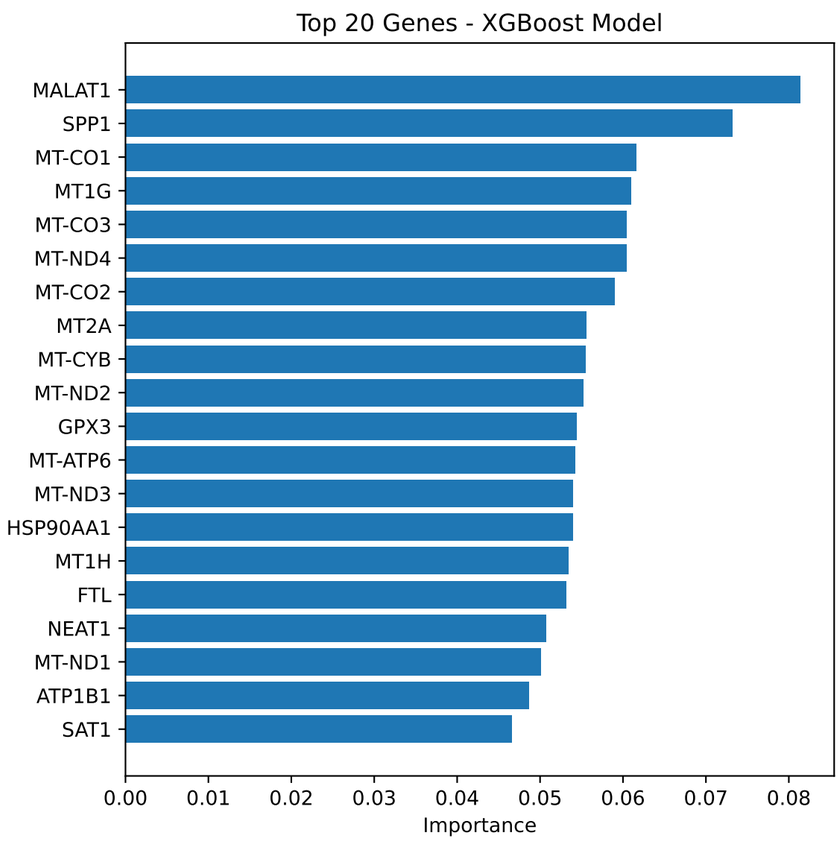
Figure S9: Evidence for exclusion of sample “kidney1” from Liao et al. Violin plot of the expression of a mitochondrial gene in each of the three samples from Liao et al showing a uniform abnormally high distribution of kidney1 (left). UMAP of cells in the “Proximal Tubule” harmonized cell type, colored by sample of origin showing that while most samples align with one another, cells from sample kidney1 appear distinct from the group and have poor alignment with the other cells in the same cell type (right).

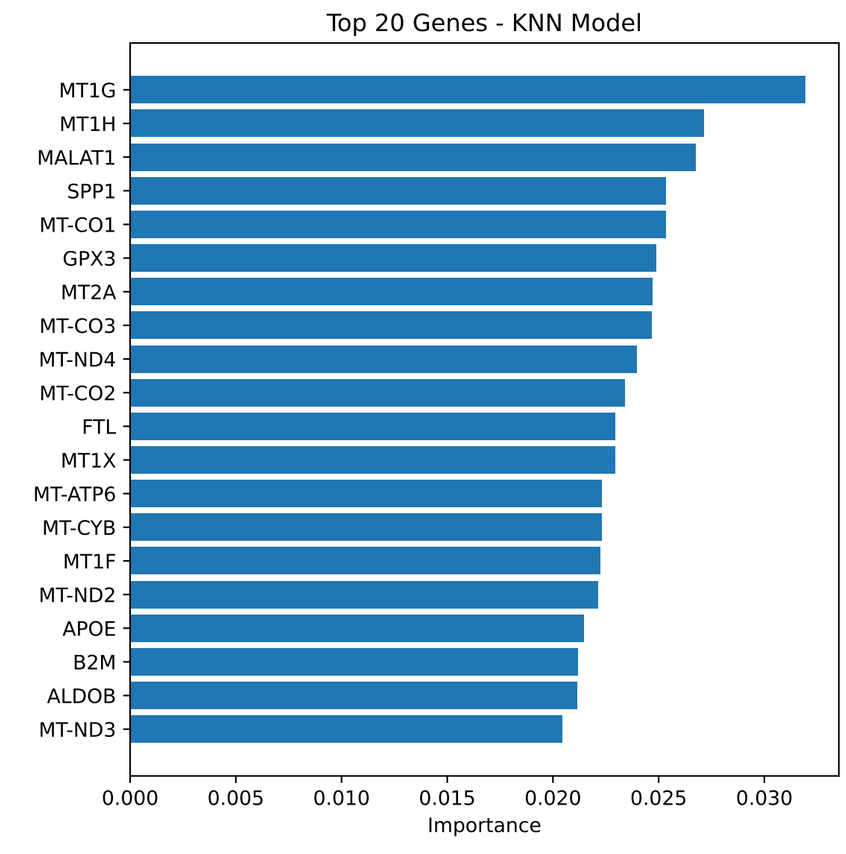



Top 20 Genes - RF Model
MALAT1
SPP1
MT-CO1
MT1G
MT-CO3
MT-ND4
HSP90AA1
MT-CO2
MT2A
MT-CYB
MT-ND2
MT1H
MT-ATP6
GPX3
NEAT1
MT-ND3
B2M
MT-ND1
ATP1B1
IGFBP7
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08
Importance


Top 20 Genes - SVC Model
ATP1B1
HSP90AA1
MALAT1
IGFBP7
NEAT1
SAT1
B2M
CD74
GPX3
SPP1
MT2A
S100A6
CRYAB
FTL
HSPA1A
HSP90AB1
ITM2B
MTRNR2L8
UGT2B7
HLA-DRA
0.00 0.02 0.04 0.06 0.08 0.10 0.12 0.14
Importance


Top 20 Genes - MLP Model
MALAT1
MT1G
SPP1
MT-CO1
MT-CO3
MT-ND4
MT-CO2
GPX3
MT1H
MT-CYB
MT-ATP6
MT-ND2
MT2A
MT-ND3
MT-ND1
FTL
MT1X
APOE
ALDOB
HSP90AA1
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07
Importance


Top 20 Genes - XGBoost Model
MALAT1
SPP1
MT-CO1
MT1G
MT-CO3
MT-ND4
MT-CO2
MT2A
MT-CYB
MT-ND2
GPX3
MT-ATP6
MT-ND3
HSP90AA1
MT1H
FTL
NEAT1
MT-ND1
ATP1B1
SAT1
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08
Importance


Top 20 Genes - KNN Model
MT1G
MT1H
MALAT1
SPP1
MT-CO1
GPX3
MT2A
MT-CO3
MT-ND4
MT-CO2
FTL
MT1X
MT-ATP6
MT-CYB
MT1F
MT-ND2
APOE
B2M
ALDOB
MT-ND3
0.000 0.005 0.010 0.015 0.020 0.025 0.030
Importance