nature methods

SUPPLEMENTARY INFORMATION

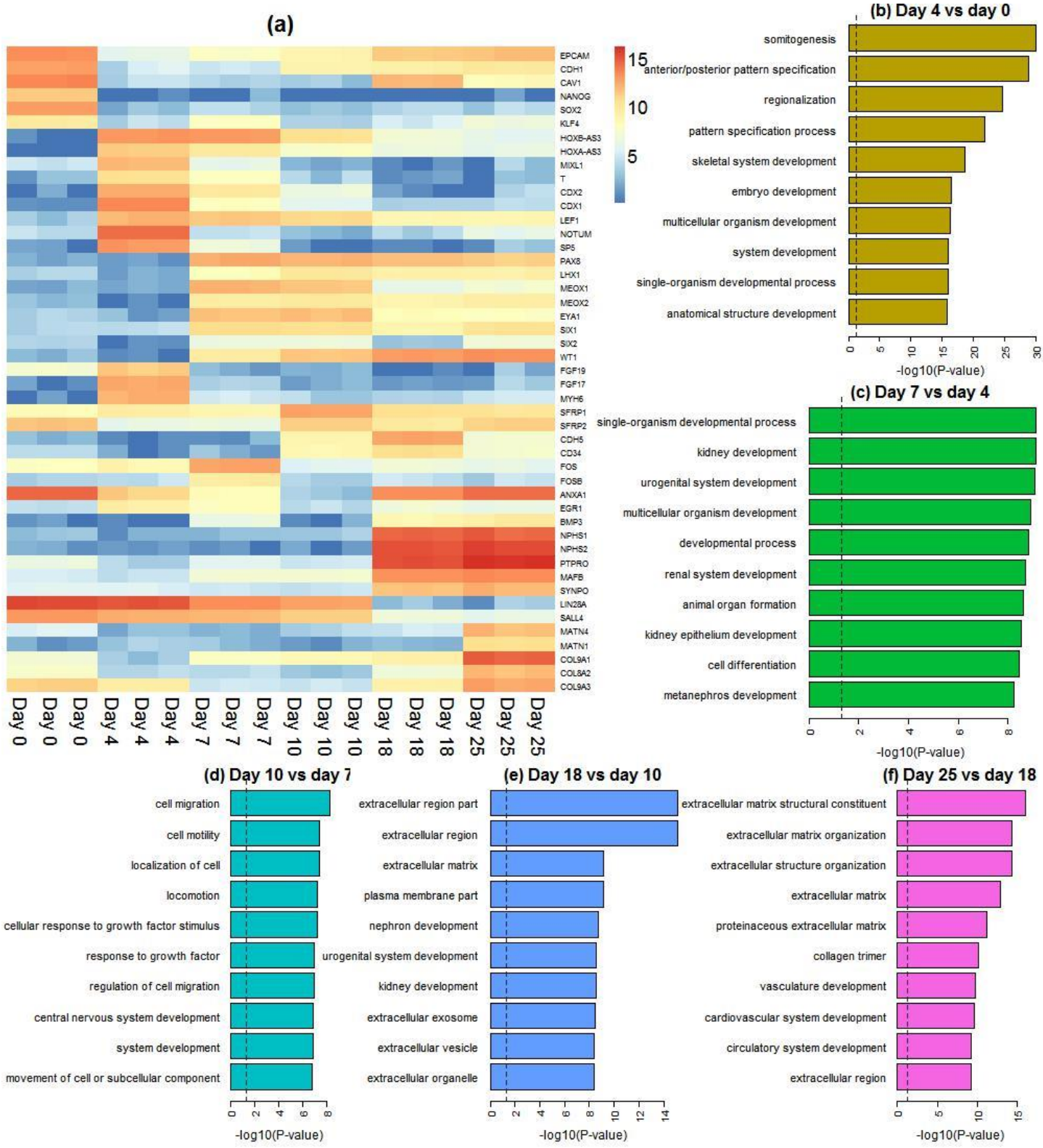
https://doi.org/10.1038/s41592-018-0253-2

In the format provided by the authors and unedited.

Evaluation of variability in human kidney organoids

Belinda Phipson[1], Pei X. Er[1], Alexander N. Combes[1,2], Thomas A. Forbes[1,3,4], Sara E. Howden[1,2], Luke Zappia[1,5], Hsan-Jan Yen[1], Kynan T. Lawlor[1], Lorna J. Hale[1,4], Jane Sun[6], Ernst Wolvetang[6], Minoru Takasato[1,7], Alicia Oshlack[1,5] and Melissa H. Little[1,2,4]*

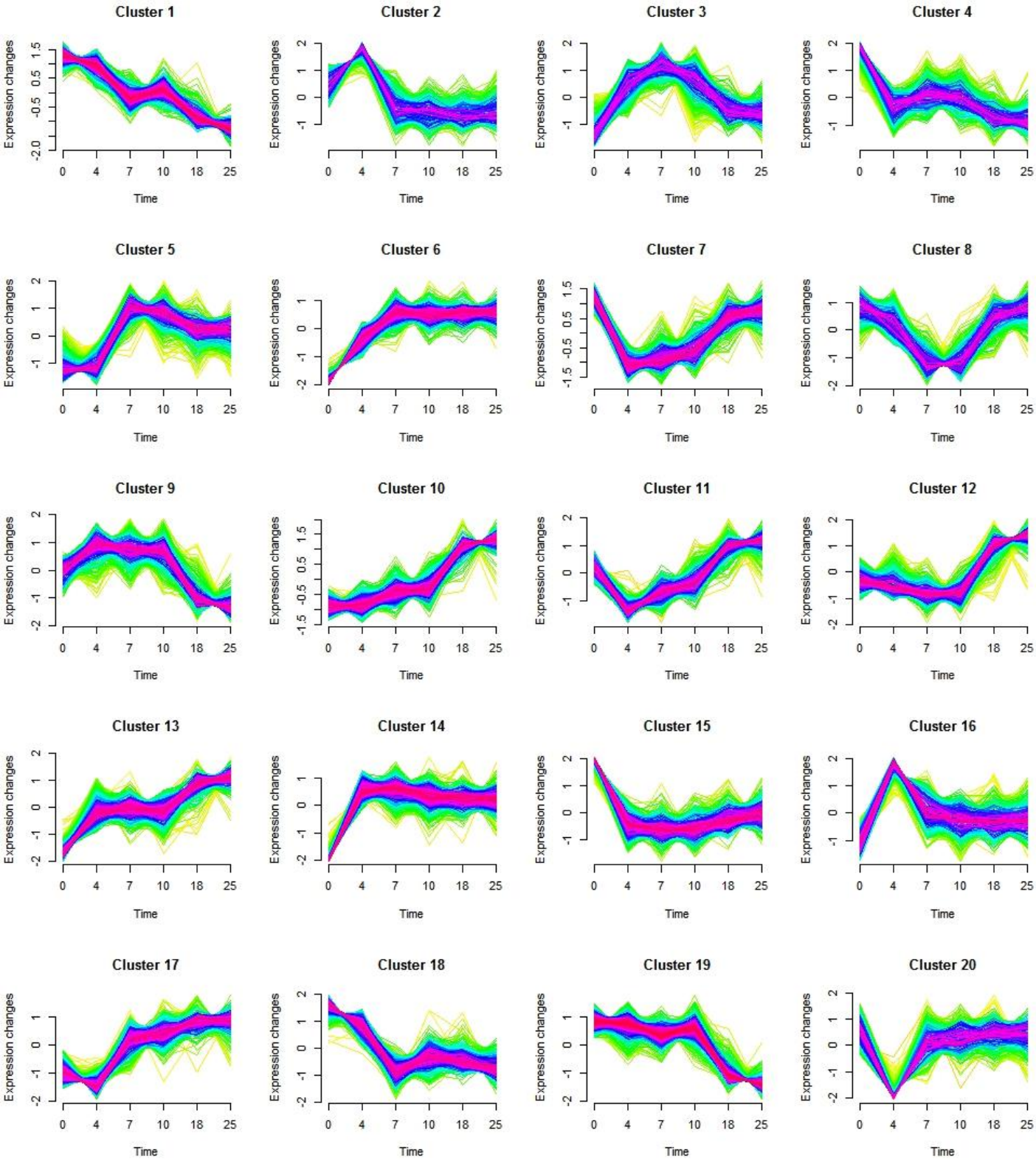
[1]Murdoch Children's Research Institute, Melbourne, Victoria, Australia. [2]Department of Anatomy and Neuroscience, The University of Melbourne, Melbourne, Victoria, Australia. [3]Department of Nephrology, Royal Children's Hospital, Melbourne, Victoria, Australia. [4]Department of Paediatrics, The University of Melbourne, Melbourne, Victoria, Australia. [5]School of Biosciences, The University of Melbourne, Melbourne, Victoria, Australia. [6]Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, Queensland, Australia. [7]Present address: RIKEN Center for Biosystems Dynamics Research, Kobe, Japan. *e-mail: melissa.little@mcri.edu.au



Supplementary Figure 1

Differentially expressed genes and enriched pathways between consecutive time points in the kidney organoid differentiation protocol.

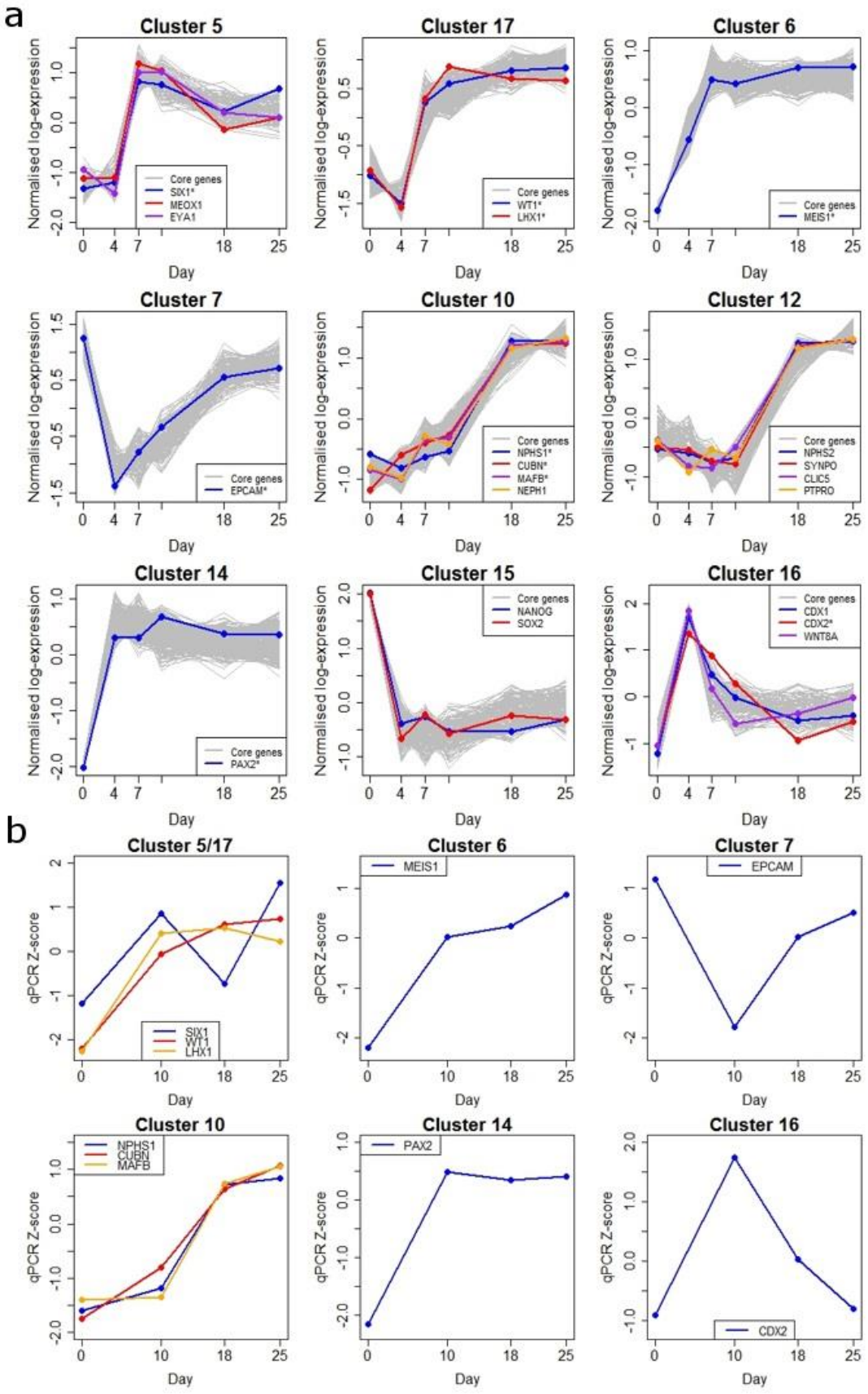
(a) Heat map of $\log_2$-normalized expression values for the top differentially expressed genes between consecutive time points in the kidney organoid differentiation protocol. Significant genes were determined with a TREAT test with an absolute log-fold-change cutoff of 1 and false discovery rate cutoff of 5% (two-sided). (b) Gene Ontology analysis of the top 100 upregulated genes between day 4 and day 0. (c) Gene Ontology analysis of the top 100 upregulated genes between day 7 and day 4. (d) Gene Ontology analysis of the top 100 upregulated genes between day 10 and day 7. (e) Gene Ontology analysis of the top 100 upregulated genes between day 18 and day 10. (f) Gene Ontology analysis of the top 100 upregulated genes between day 25 and day 18. The top 10 gene ontology categories for each comparison are shown. The top 100 upregulated genes between each comparison were tested for enrichment in the GO categories. A modified hypergeometric test was used to determine statistical significance, taking into account gene length bias. *P* values are one-sided.



Supplementary Figure 2

Expression patterns for 20 fuzzy clusters identified across the time points of the kidney organoids differentiation protocol.

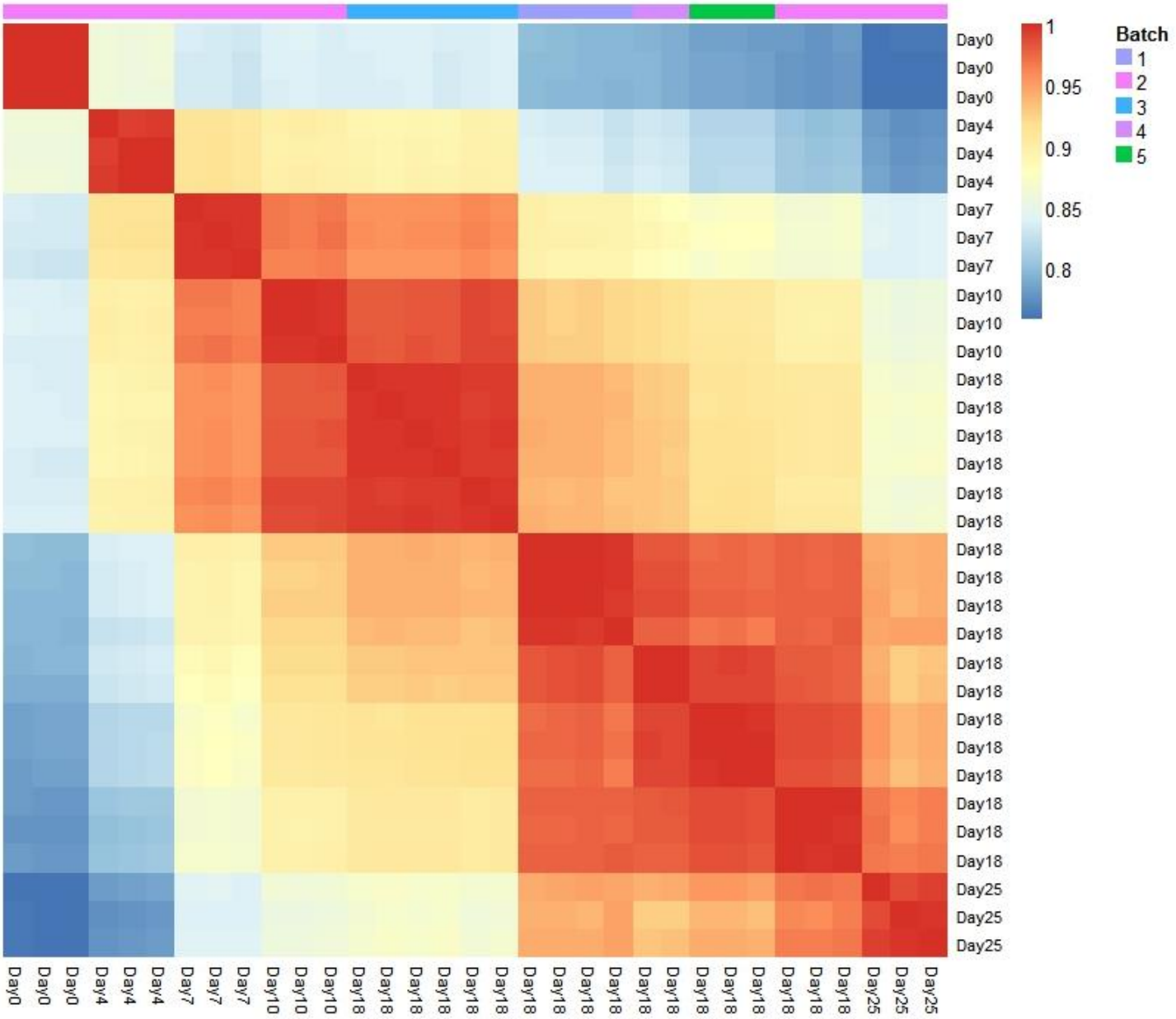
Genes that displayed similar patterns of expression across the time course data were clustered using fuzzy c-means clustering, a soft clustering approach that assigns each gene gradual degrees of membership to each of the 20 clusters. We identified 7,682 genes to use as input for the Mfuzz algorithm. These genes were changing between at least one time point in the differentiation protocol.


a
Cluster 5
Core genes
SIX1*
MEOX1
EYA1
Cluster 17
Core genes
WT1*
LHX1*
Cluster 6
Core genes
MEIS1*
Cluster 7
Core genes
EPCAM*
Cluster 10
Core genes
NPHS1*
CUBN*
MAFB*
NEPH1
Cluster 12
Core genes
NPHS2
SYNPO
CLIC5
PTPRO
Cluster 14
Core genes
PAX2*
Cluster 15
Core genes
NANOG
SOX2
Cluster 16
Core genes
CDX1
CDX2*
WNT8A
Normalised log-expression
Day
b
Cluster 5/17
SIX1
WT1
LHX1
Cluster 6
MEIS1
Cluster 7
EPCAM
Cluster 10
NPHS1
CUBN
MAFB
Cluster 14
PAX2
Cluster 16
CDX2
qPCR Z-score
Day

Supplementary Figure 3

Analysis of the molecular program of human kidney differentiation.

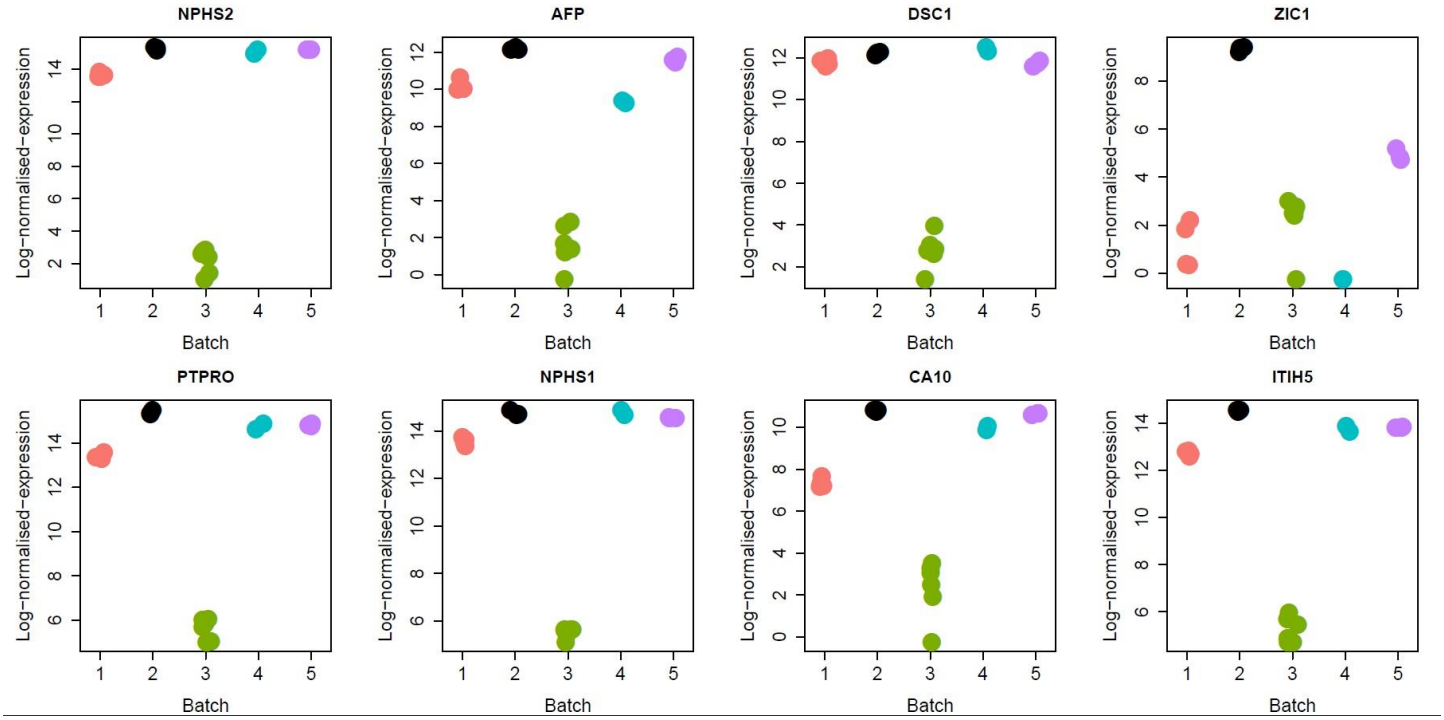
(a) Selected cluster plots showing synexpression across the developmental time. Core genes with membership scores > 0.5 are shown in gray. Example genes of interest are shown in color for each cluster. Genes with an asterisk were validated with qPCR. Lists of core genes for each cluster are seen in Supplementary Table 3. (b) qPCR validation of selected genes from fuzzy clusters. At each time point qPCR was performed to measure expression of 10 genes in biologically independent samples. Each time point has $n = 1$ independent replicates ($n = 4$ in total).



Supplementary Figure 4

Pairwise correlation coefficients between all CRL1502-C32 organoids.

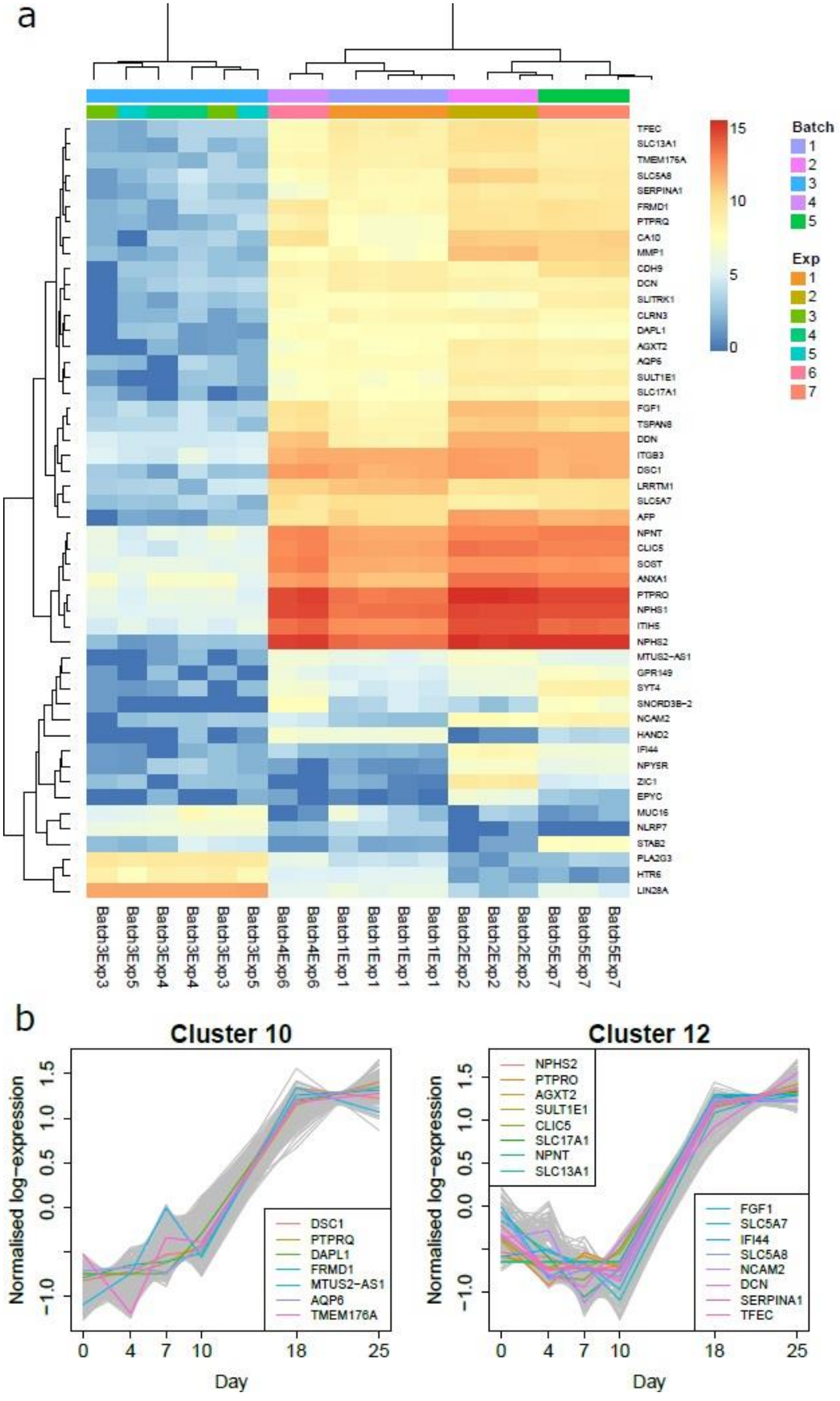
Spearman's rho statistic was used to provide a rank-based measure of association. Correlations are estimated from 15,685 genes.



Supplementary Figure 5

Top eight most variable genes, stratified by batch.

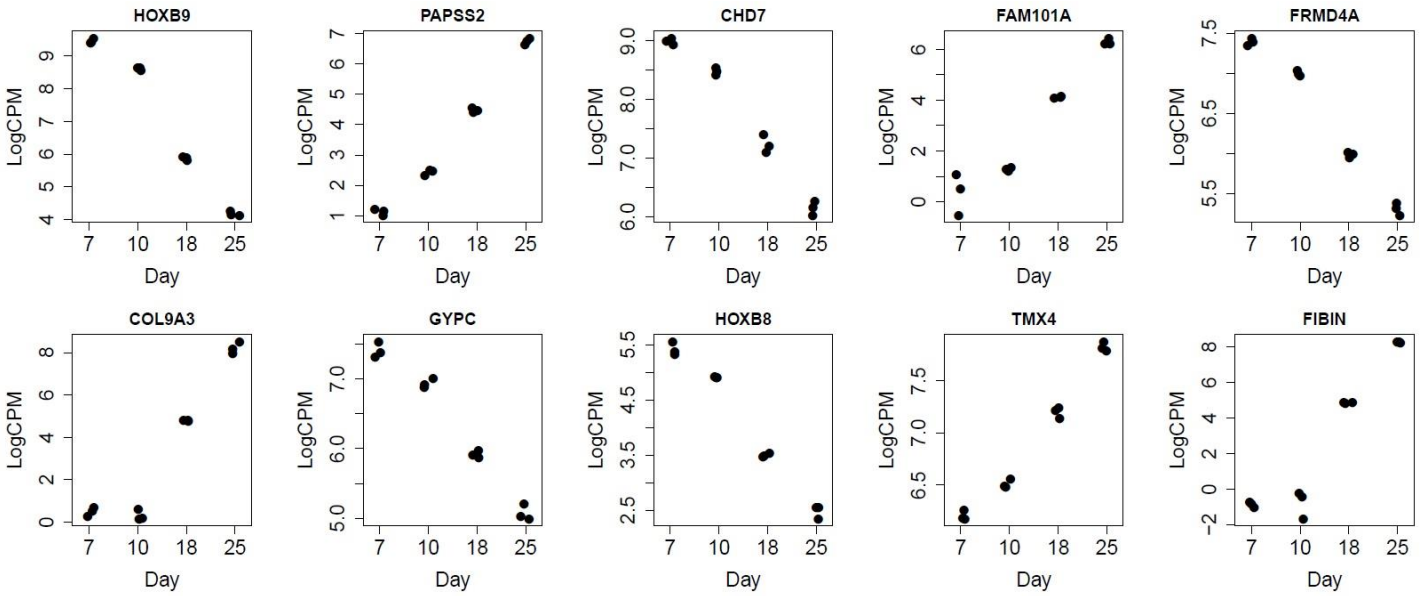
Points are colored according to batch as in Fig. 1b. Supplementary Fig. 6a shows a heat map of the top 50 most variable genes, and a full table of variability analysis results is presented in Supplementary Table 6.


a
TFEC
SLC13A1
TMEM176A
SLC5A8
SERPINA1
FRMD1
PTPRQ
CA10
MMP1
CDH9
DCN
SLITRK1
CLRN3
DAPL1
AGXT2
AQP6
SULT1E1
SLC17A1
FGF1
TSPAN8
DDN
ITGB3
DSC1
LRRTM1
SLC5A7
AFP
NPNT
CLIC5
SOST
ANXA1
PTPRO
NPHS1
ITIH5
NPHS2
MTUS2−AS1
GPR149
SYT4
SNORD3B−2
NCAM2
HAND2
IFI44
NPY5R
ZIC1
EPYC
MUC16
NLRP7
STAB2
PLA2G3
HTR6
LIN28A
Batch3Exp3
Batch3Exp5
Batch3Exp4
Batch3Exp4
Batch3Exp3
Batch3Exp5
Batch4Exp6
Batch4Exp6
Batch1Exp1
Batch1Exp1
Batch1Exp1
Batch1Exp1
Batch2Exp2
Batch2Exp2
Batch2Exp2
Batch5Exp7
Batch5Exp7
Batch5Exp7
15
10
5
0
Batch
1
2
3
4
5
Exp
1
2
3
4
5
6
7
b
Cluster 10
Normalised log−expression
−1.0
0.0
0.5
1.0
1.5
DSC1
PTPRQ
DAPL1
FRMD1
MTUS2−AS1
AQP6
TMEM176A
0
4
7
10
18
25
Day
Cluster 12
NPHS2
PTPRO
AGXT2
SULT1E1
CLIC5
SLC17A1
NPNT
SLC13A1
FGF1
SLC5A7
IFI44
SLC5A8
NCAM2
DCN
SERPINA1
TFEC

Supplementary Figure 6

Highly variable genes between CRL1502-C32 day 18 organoids.

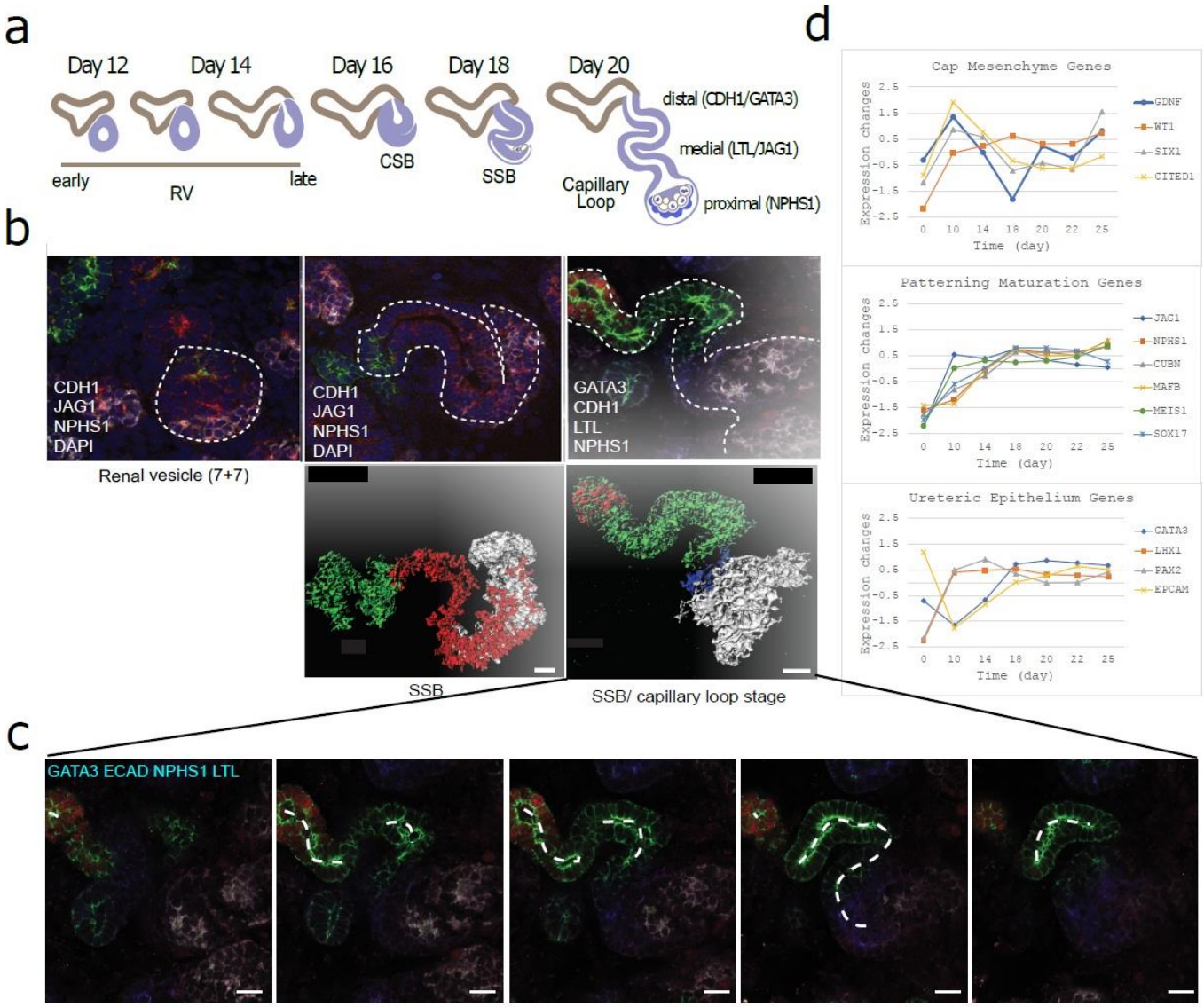
(a) Unsupervised hierarchical clustering shows that organoids within batches cluster together. Expression values are $\log_2$-normalized. (b) Highly variable genes present in clusters 10 and 12. The overlap between the top 50 most variable genes and core genes in each cluster showed 7 and 16 genes in common in clusters 10 and 12, respectively. These genes are shown in color.



Supplementary Figure 7

Identification of key genes that predict relative maturation.

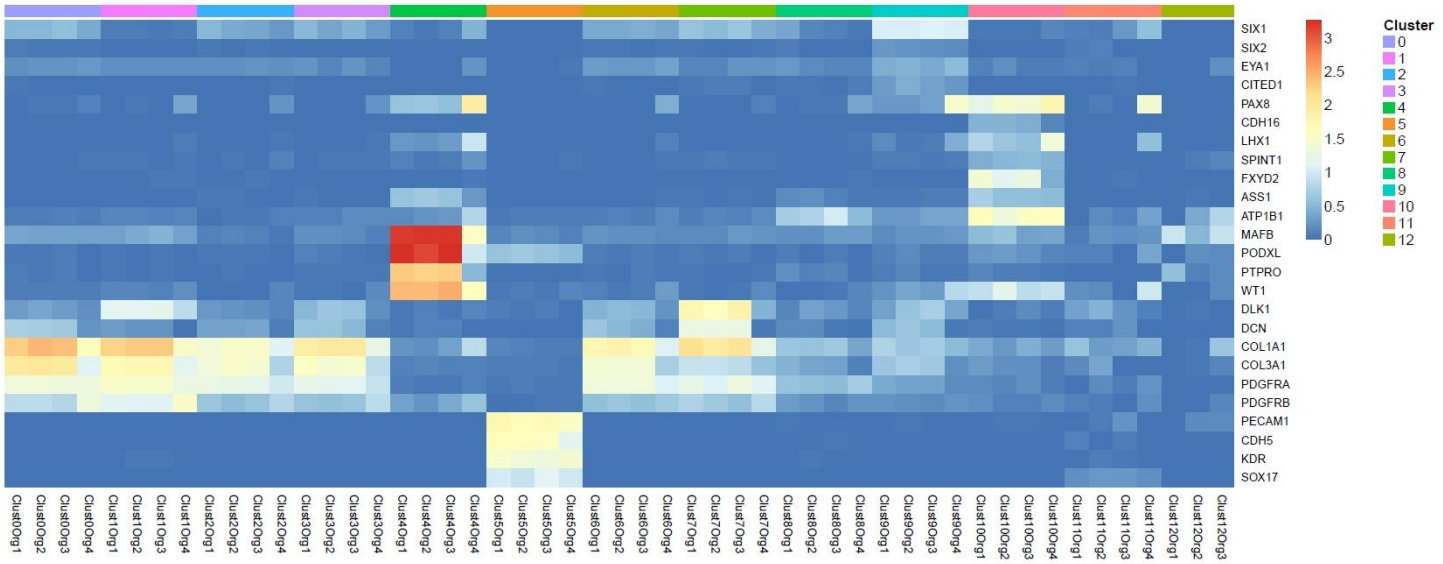
The 10 genes most significantly linearly related with the time series between days 7 and 25 were selected to build a multivariate linear model to predict organoid maturity of new samples.



Supplementary Figure 8

Evidence for nephron patterning and segmentation to capillary loop stage.

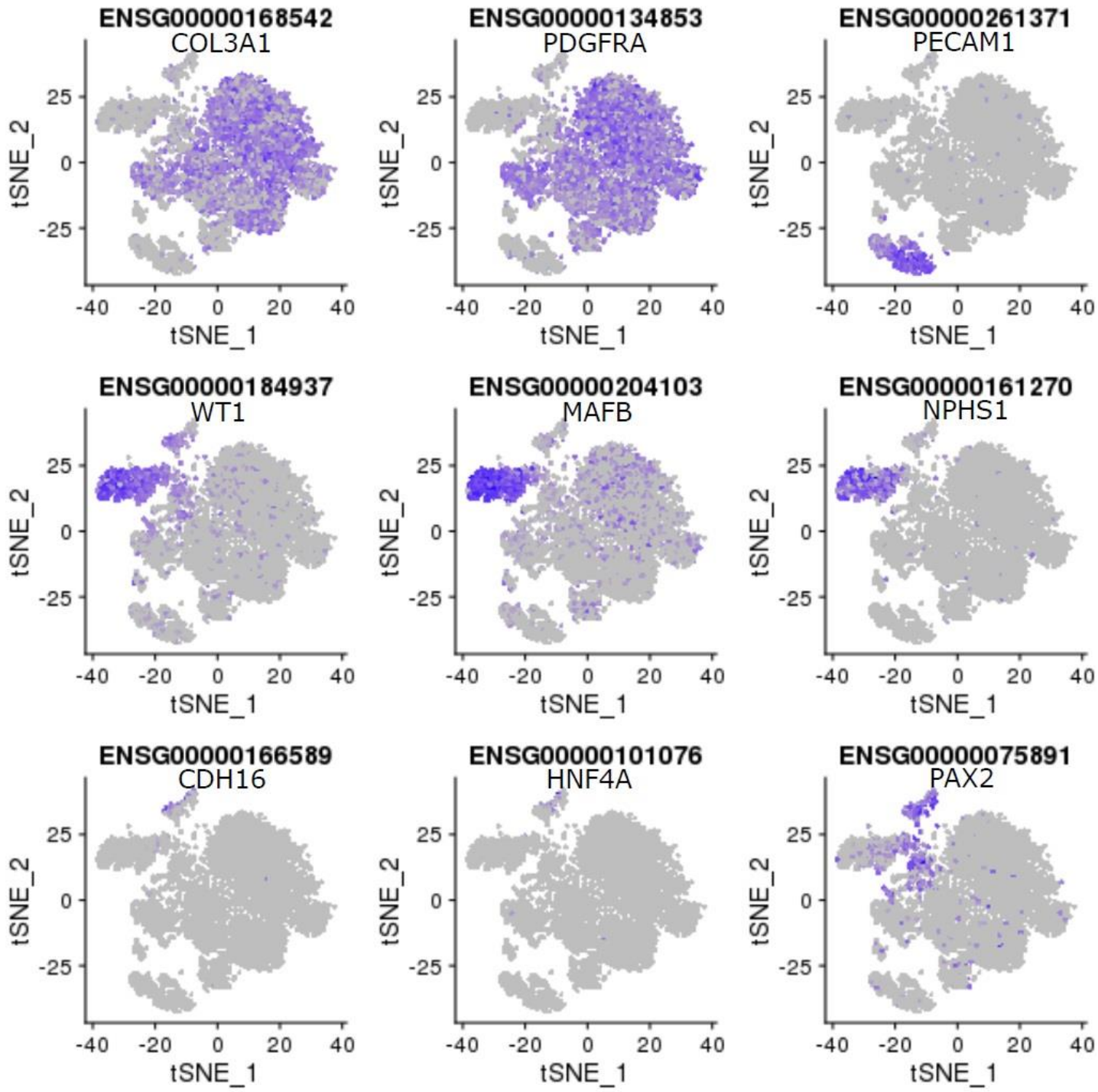
a. Diagram of anticipated morphological changes across nephron formation in mouse kidney. **b**. Immunofluorescence of nephrons forming within organoids across time using markers for proximal nephron (white; NPHS1), distal RV / medial nephron (red; JAG1) and distal nephron (green; CDH1). **c.** Serial single Z slices through confocal images of a single capillary-loop stage nephron showing segmentation along the length of the tubule and evidence for a lumen passing in and out of the plane of the image. Immunofluorescence performed using antibodies to markers of collecting duct (red; GATA3), epithelium (green; CDH1) and proximal nephron (white; NPHS1). Scale bar, 10 μm. **d.** QPCR of key nephrogenic genes between days 10 and 25 shows gradual loss of nephrogenic progenitors from day 10 to day 14 and formation of nephrons commencing at the same time. All images are representative images observed from within at least three organoids in this particular experiment. For the qPCR, each time point for each gene represents experimental triplicates.



Supplementary Figure 9

Heat map of log-normalized expression of selected kidney marker genes in single-cell data.

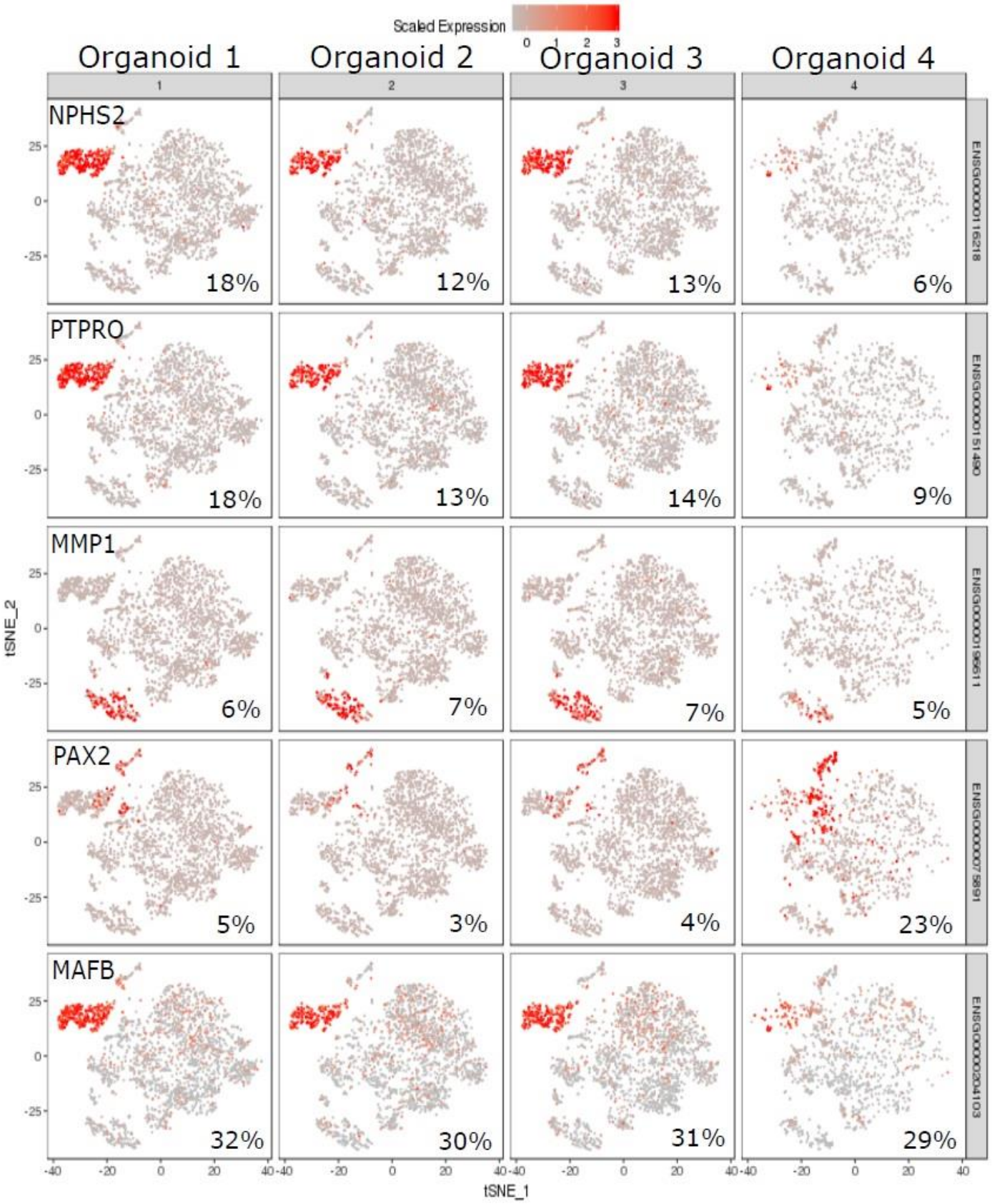
Expression values have been averaged across the cells per organoids per cluster.



Supplementary Figure 10

tSNE plots showing expression and distribution of selected kidney marker genes.

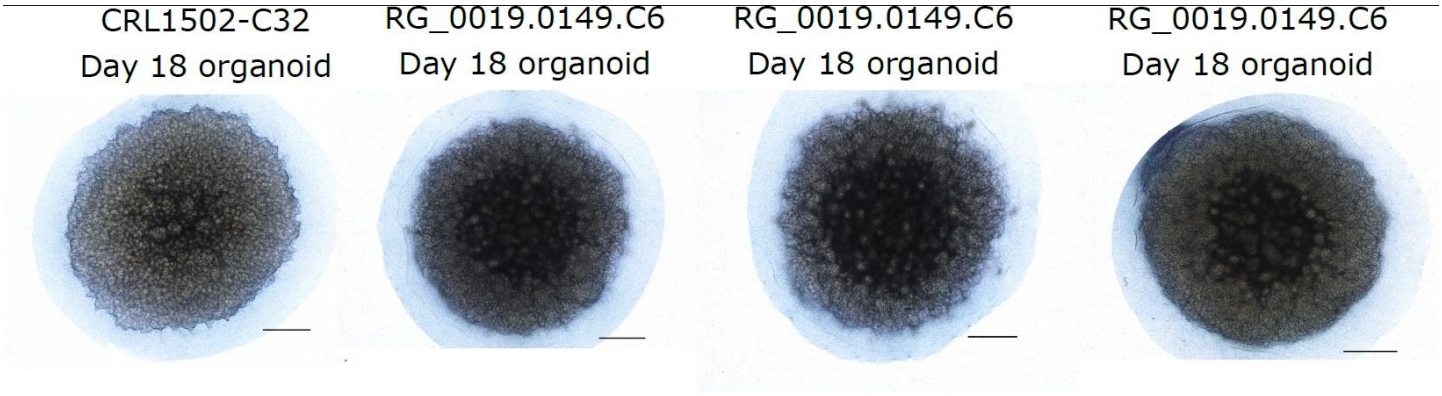
Each tSNE is made up of 8,361 cells from *n* = 4 biologically independent organoids.


Scaled Expression
0 1 2 3
Organoid 1
Organoid 2
Organoid 3
Organoid 4
1
2
3
4
NPHS2
PTPRO
MMP1
PAX2
MAFB
18%
12%
13%
6%
18%
13%
14%
9%
6%
7%
7%
5%
5%
3%
4%
23%
32%
30%
31%
29%
ENSG00000116218
ENSG00000151490
ENSG00000196611
ENSG00000075891
ENSG00000204103
tSNE_2
tSNE_1
25
0
-25
-40
-20
0
20
40

Supplementary Figure 11

Stratified tSNE plot (by organoid) showing expression and distribution of three variable genes (*NPHS2*, *PTPRO* and *MMP1*) and two kidney marker genes (*PAX2* and *MAFB*).

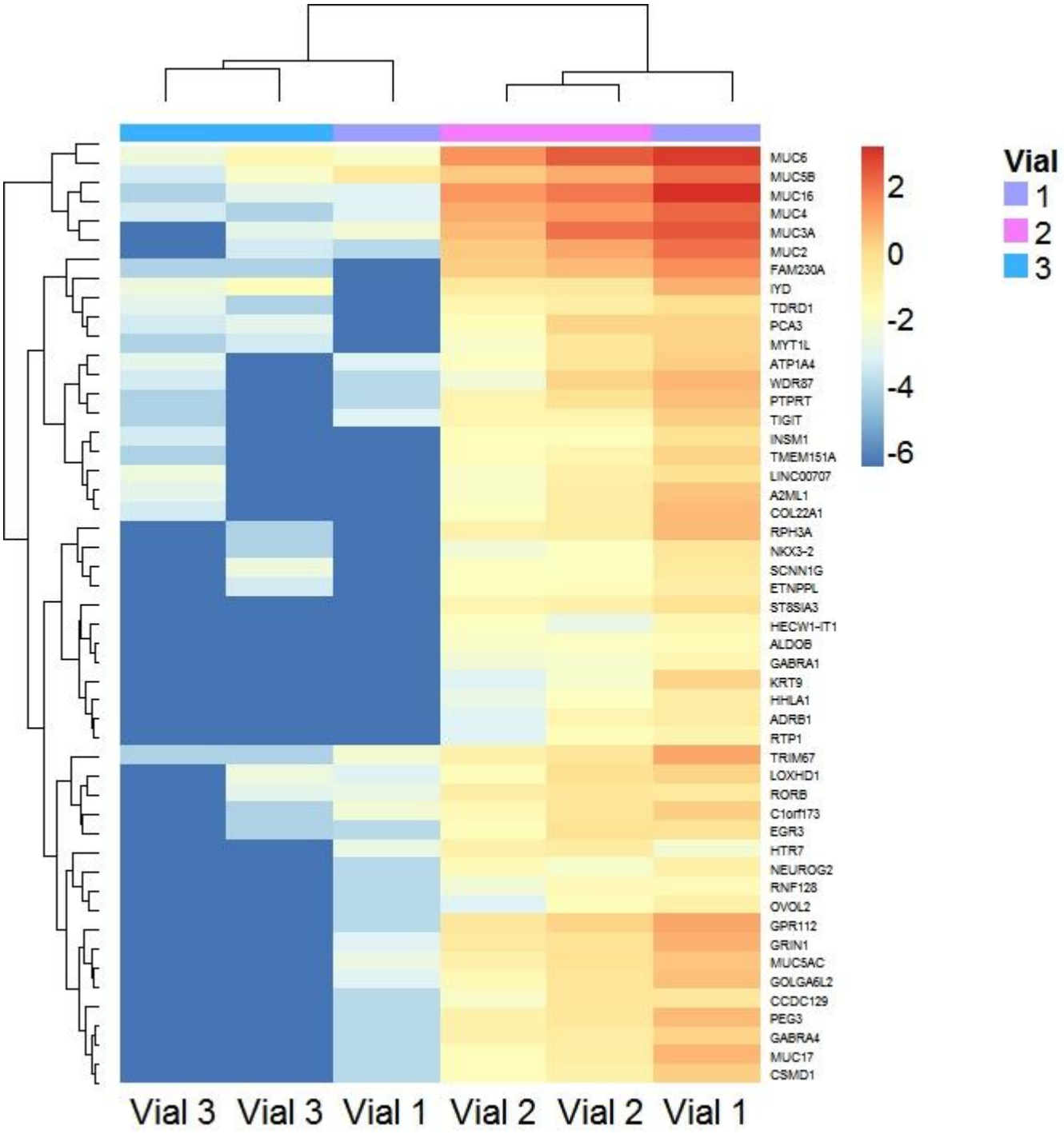
Each tSNE has 2,414 cells for organoid one, 2,202 cells for organoid two, 2,289 cells for organoid three and 1,418 cells for organoid four.



Supplementary Figure 12

Bright-field images of cell line RG_0019.0149.C6 displaying similar nephron formation and segmentation compared to the CRL1502-C32 line.

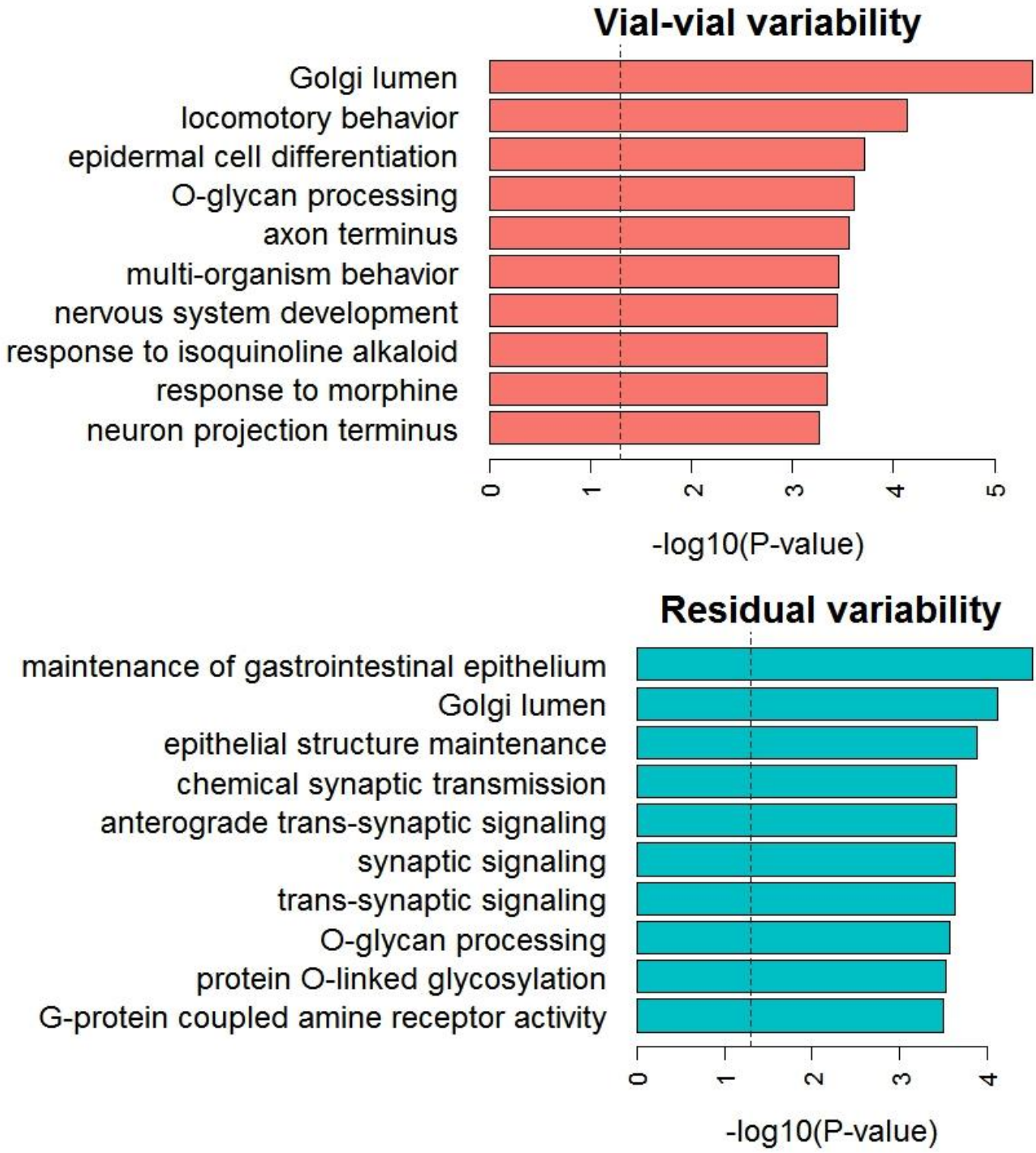
Scale bar is representative of 1 mm. There are three biologically independent organoids shown here for the RG_0019.0149.C6 line and one biologically independent organoid shown for CRL1502-C32.



Supplementary Figure 13

Heat map for the top 50 most highly variable genes between RG_0019.0149.C6 day 18 organoids.

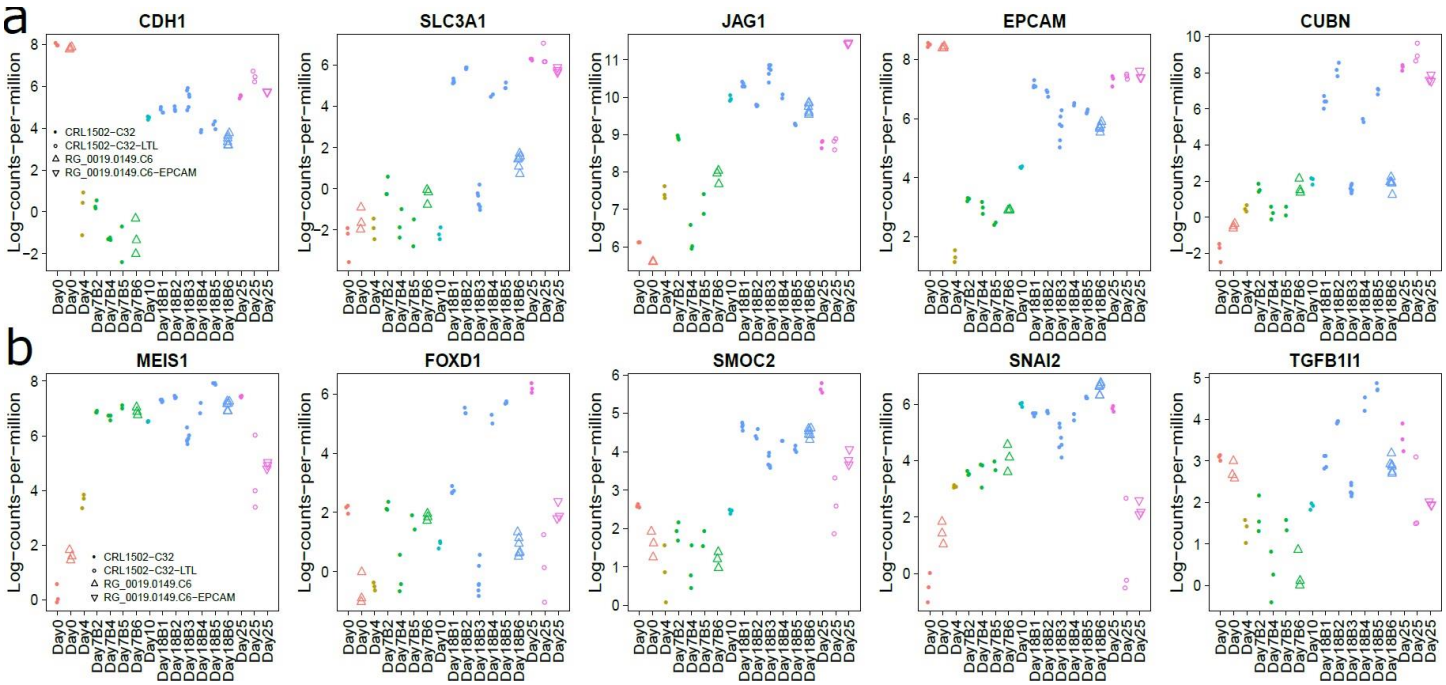
Expression values are $\log_2$-normalized. Highly variable genes were identified using a random effects model.



Supplementary Figure 14

Top 10 enriched GO terms for genes contributing to vial-to-vial and residual variability between RG_0019.0149.C6 organoids.

The top 100 most variable genes were tested for enrichment in the GO categories. A modified hypergeometric test was used to determine statistical significance, taking into account gene length bias. *P* values are one-sided. GO categories with at least 10 genes are shown.



Supplementary Figure 15

Expression of key epithelial and interstitial genes across all organoids.

(a) Log-normalized expression of five epithelial-related genes for all total organoids and enriched nephron epithelium samples. (b) Log-normalized expression of five interstitial-related genes for all total organoids and enriched nephron epithelium samples. Day 7 and day 18 samples are stratified by batch.