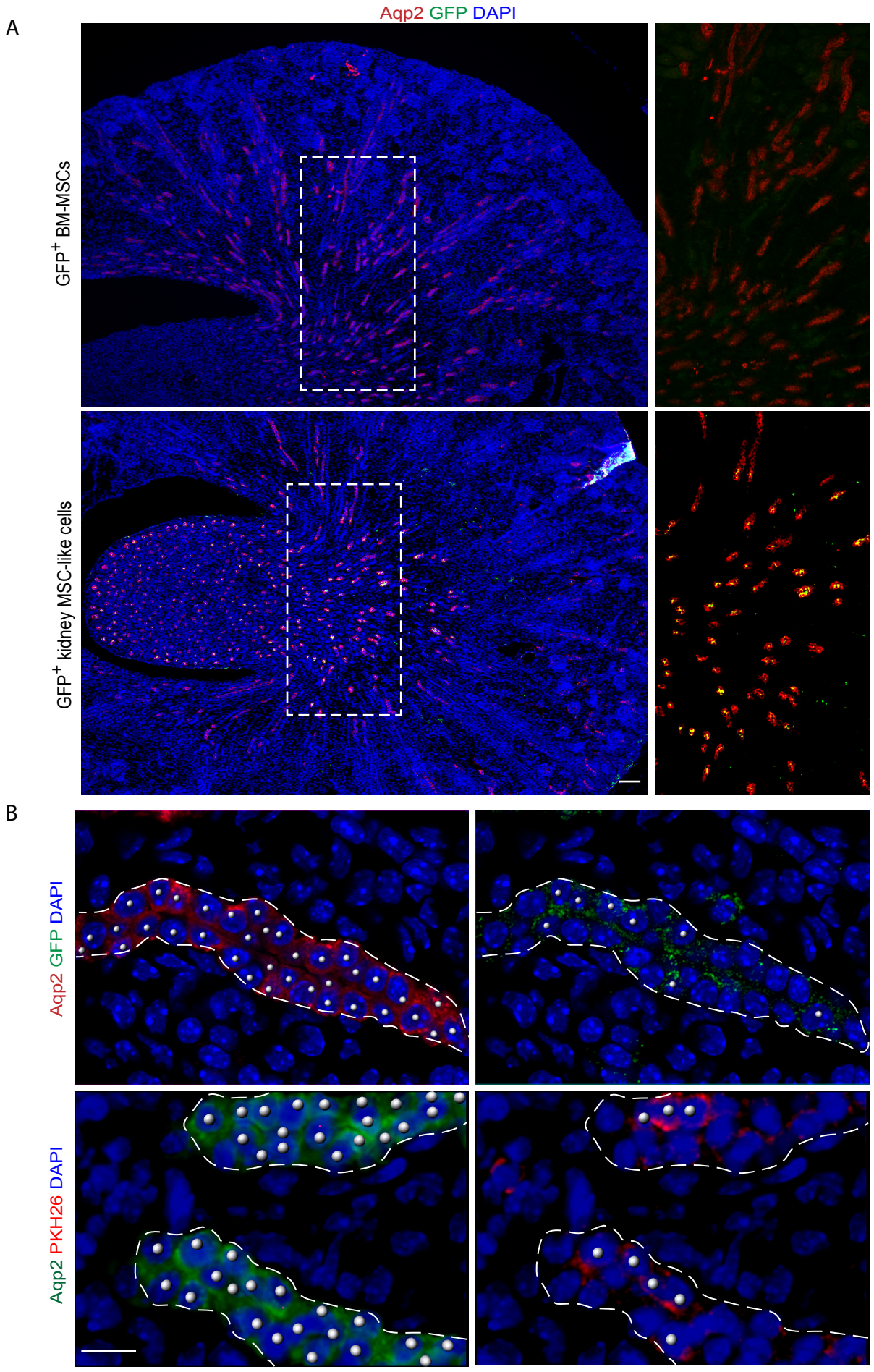
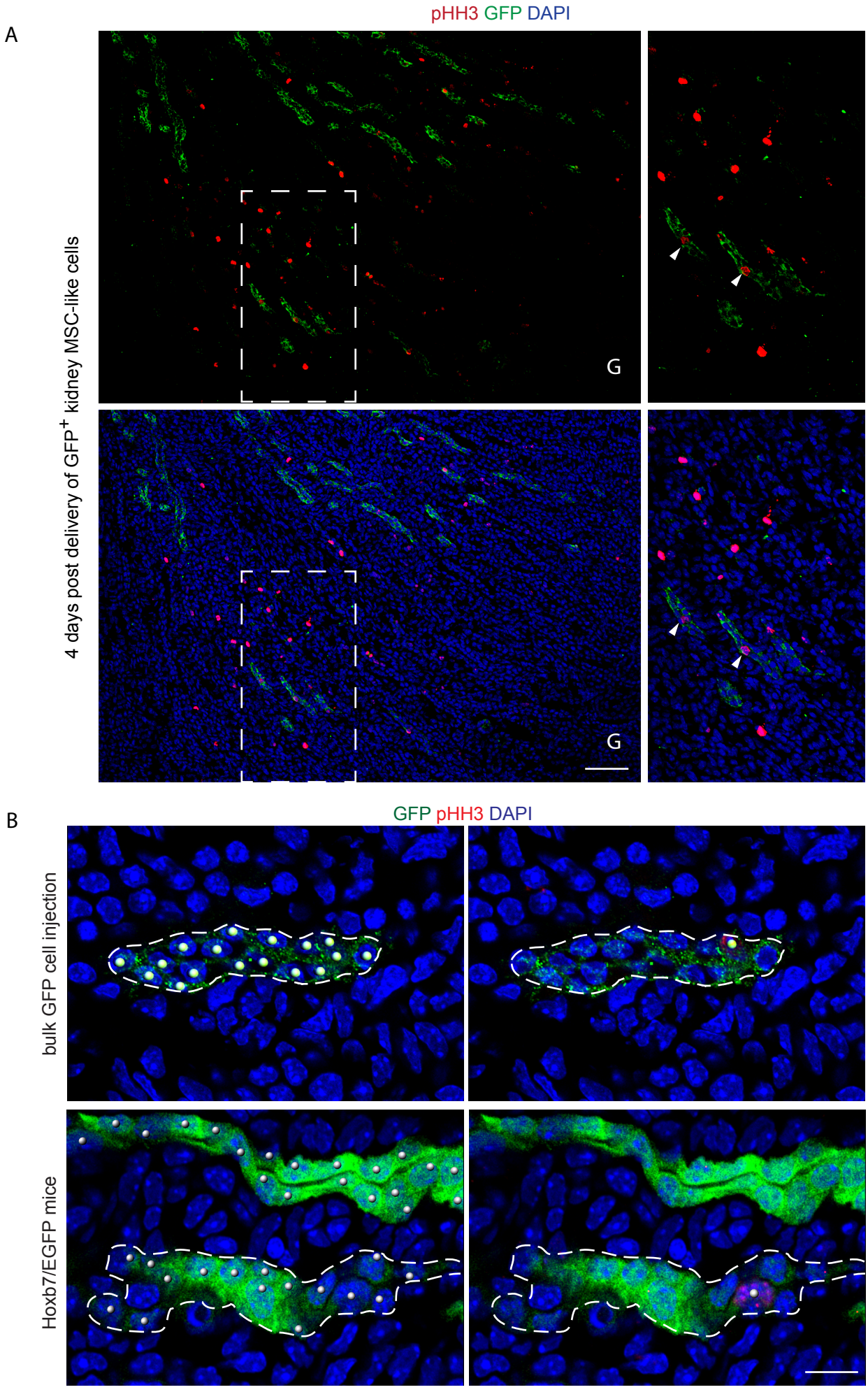
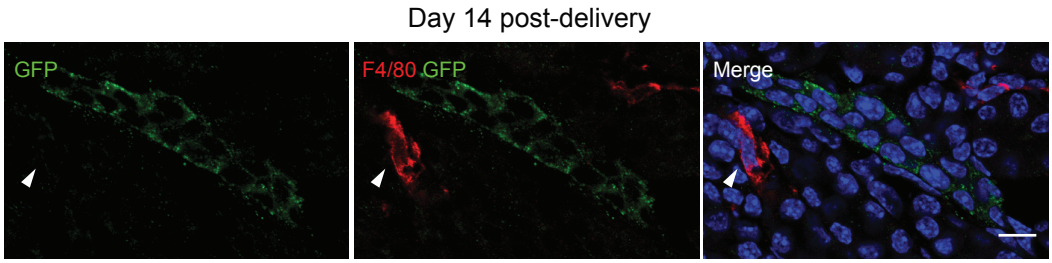
Supplementary Figure 1



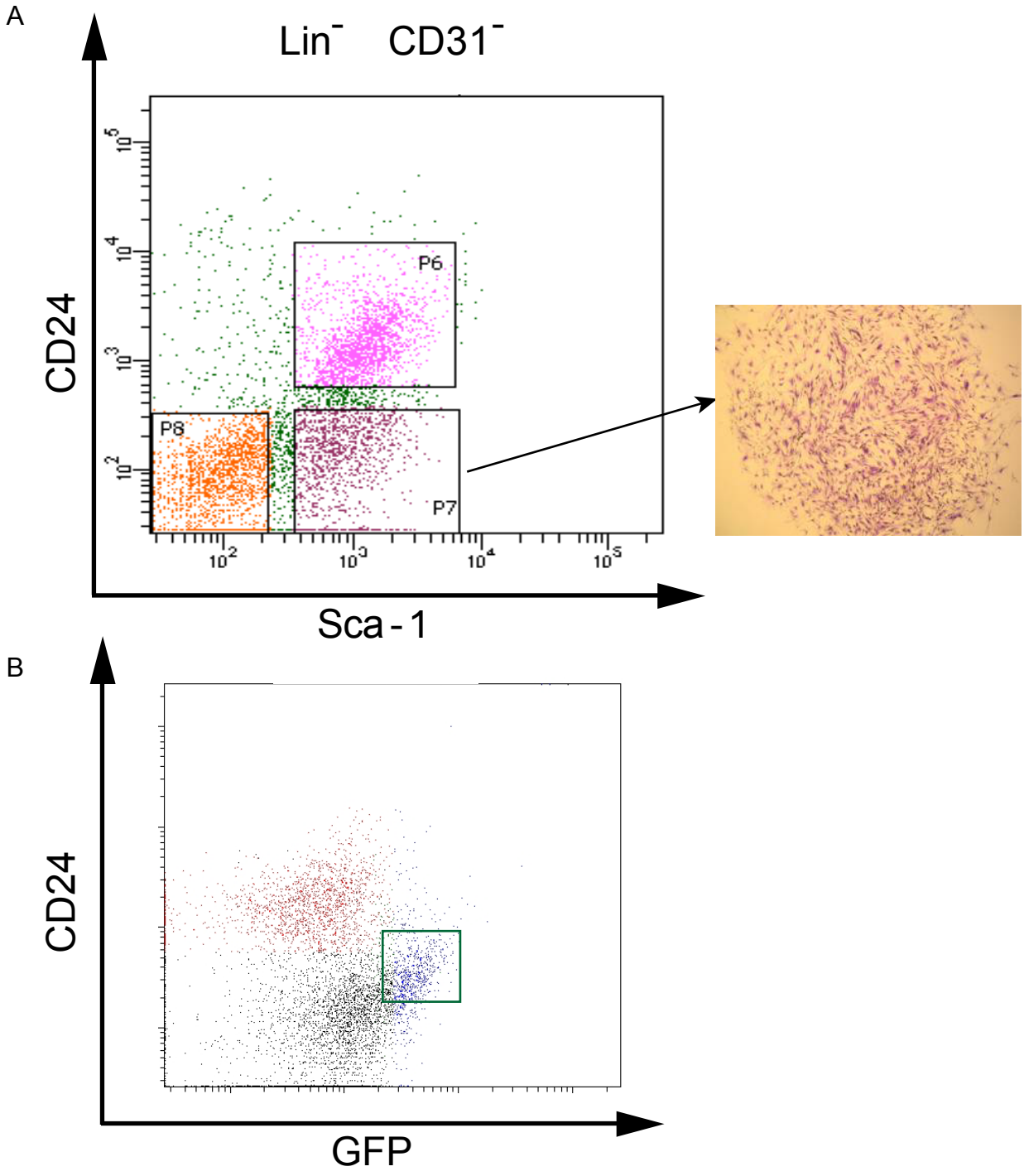
Supplementary Figure 2



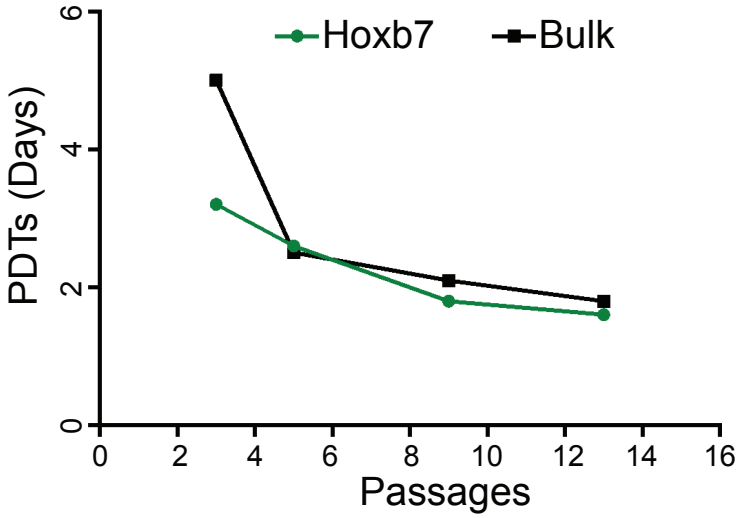
Supplementary Figure 3



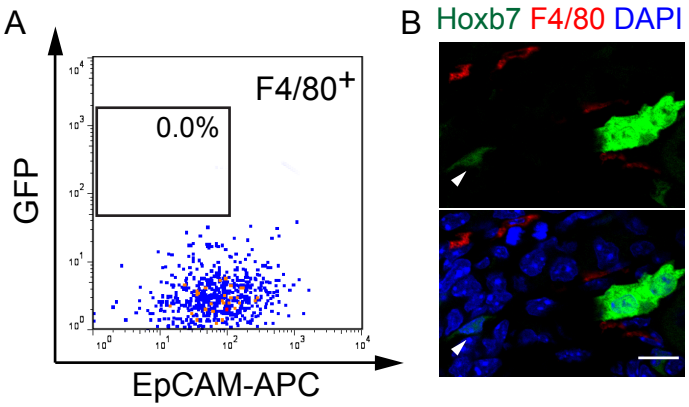
Supplenmentary Figure 4



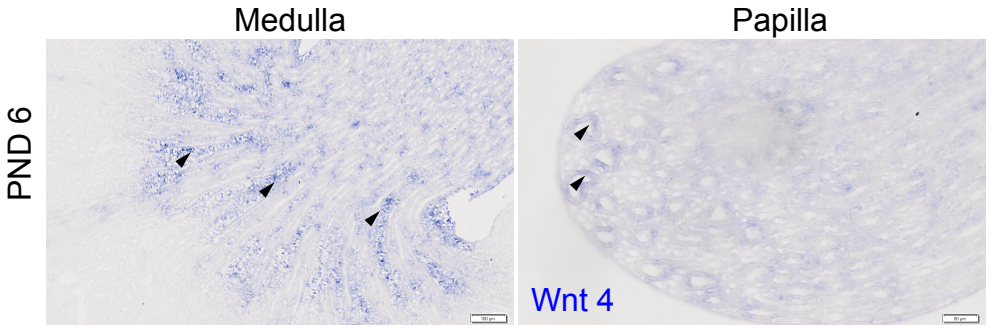
Supplementary Figure 5



Supplementary Figure 6

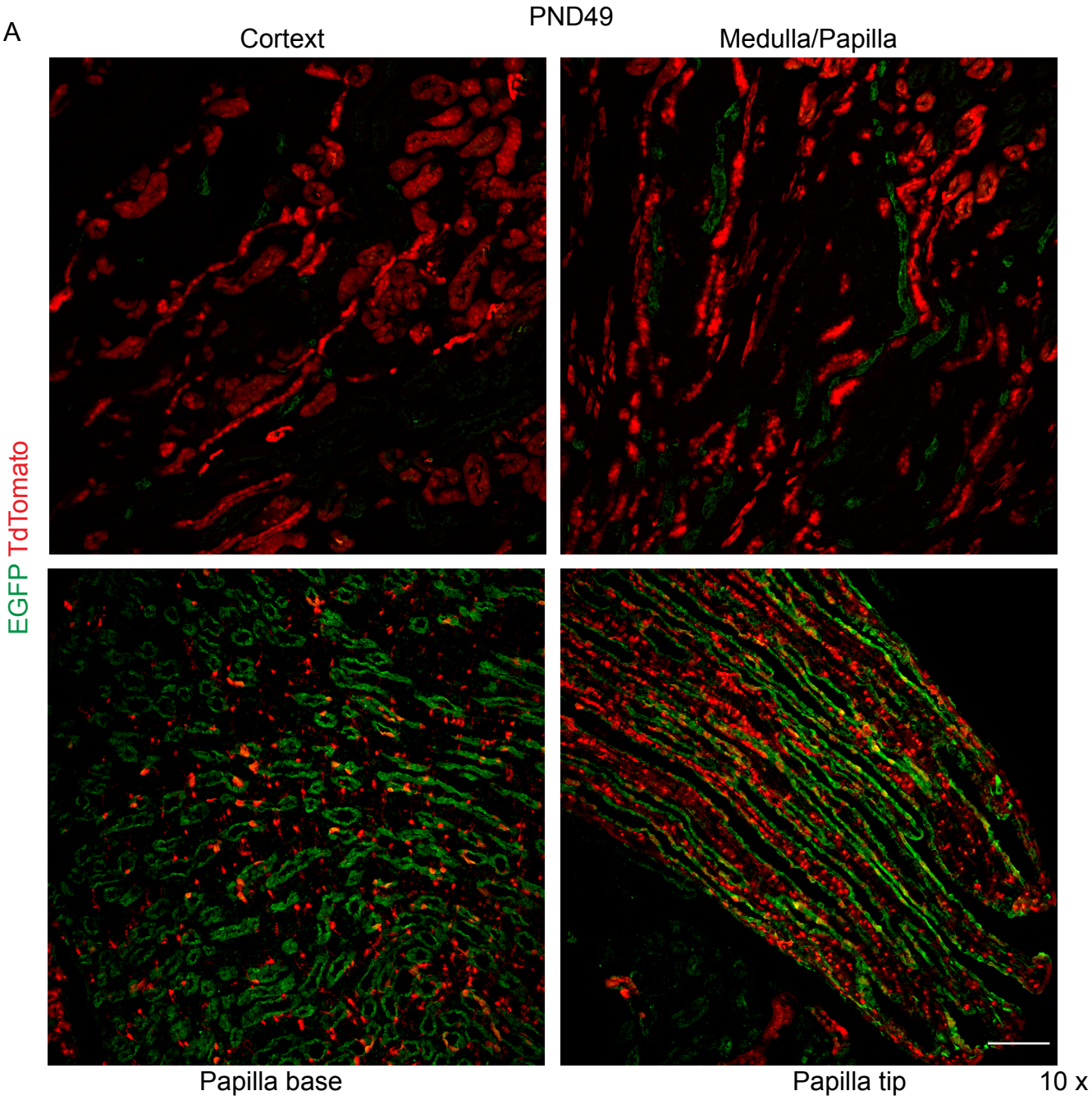


Supplementary Figure 7



Supplementary Figure 8

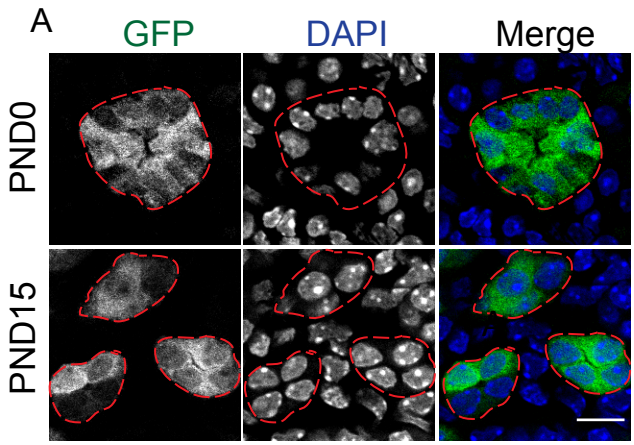
A



B

	Collecting duct Aqp2^{+}DAPI^{+}				Wnt4-expressing EGFP^{+}DAPI^{+}		
	EGFP^{+}	TdTomato^{+}	double	total	Apq2^{-}	TdTomato^{+}	total
E19.5	0	0	0	813	371	31	371
PND49	878	73	73	878	0	73	878

Supplementary Figure 9



Supplementary Table 1

qPCR primer sequences

Gene	Forward Primer (5’ – 3’)	Reverse Primer (5’ – 3’)
GAPDH	CCCAATGTGTCCGTCGTG	GCCTGCTTCACCACCTTCT
NG2	CCCGATGATGTAGGTGATGCTTC	GCTGCCCTGTAGTGAAACACAAC
Pod1	CCCACTAAGAAAAGCCCGCTC	CCGTTCTCGTACTTGTCGTTG
Wnt4	AGTGGAGAACTGGAGAAGTGT	CAAAGGACTGTGAGAAGGCTA
Ecad	ATACACTCTGGTGGTTCAGGCTGC	CCTCATTCTCAGGCACTTGACCC
Hoxb7	CTGGCGCCAAGGAGCAGAGG	CGGCCTCGCTTTCGGTCAGG
Wnt9b	GCAGGAGCGCTGGAACTGCA	CCAGTGCATGCGTGAGGGCA
mTert	AGCAAAAACCTTCCTCAGCA	AGTGAGCAGGCAGCTGGTAT

Supplementary Figure legends:

Supplementary Figure 1. A. Low resolution confocal images show the GFP^+ tubular structures in the medulla region of neonatal kidney 4 days after injection. Scale bar: 100 μm. **B.** High resolution confocal images demonstrate the masking and spot counting strategy performed using Imaris software for quantification of integration capacity of bulk GFP^+ cells or PKH-26 labelled Hoxb7-derived cells. Quantification on the integration efficiency was assessed using Imaris software (Version 7.2, BitPlane AG). Separate spot counting was performed to determine the number of $DAPI^+$ cells within the $Aqp2^+$ structures and the number of $DAPI^+$ nuclei within the GFP^+ (or $PKH26^+$) structures. Co-localisation of Aqp2 and GFP (or PKH26) was manually checked to ensure every GFP^+ (or $PKH26^+$) cell counted was also positive for Aqp2 staining. A total of 6288 $Aqp2^+$ cells were counted and 12±0.8% of these cells were positive for GFP. PKH26-labelled Hoxb7-derived cells accounted for 19±3.2% of total 3144 $Aqp2^+$ cells in the chosen regions. Scale bar 10μm.

Supplementary Figure 2. A. Low resolution confocal images show the localisation of GFP^+ tubular structures in the medulla region and detection of proliferation based on positive staining for mitotic marker pHH3 (red) 4 days after injection. Arrowheads indicate occasional $pHH3^+$ cells within the GFP^+ structures. Scale bar 100μm. **B**. High resolution confocal images demonstrate the masking and spot counting strategy performed using Imaris software for quantification of proliferation rate of integrated GFP^+ kidney MSCs and native collecting duct epithelial cells (Hoxb7/$EGFP^+$) in postnatal day 6 kidney. A total of 4780 GFP^+DAPI^+ nuclei were counted using Imaris and 1.2±0.2% of these nuclei were double positive for pHH3 and GFP. Analysis for normal proliferation rate of collecting duct epithelial cells using the Hoxb7/EGFP mice showed that 0.5±0.2% (total 3044 nuclei counted) of Hoxb7/$EGFP^+$

cells were pHH3^{+} positive in a normal postnatal day 6 kidney (comparable age). Scale bar 10µm.

Supplementary Figure 3. No overlap between macrophages and GFP after neonatal injection of kidney derived-stromal cells. Immunofluorescence staining for the macrophage marker F4/80 (red), GFP (green) and DAPI (blue) in kidney sections at day 14 post-delivery of GFP^{+} kidney-derived stromal cells. Arrowhead indicates F4/80^{+} cells close to a GFP^{+} tubular structure. Scale bar: 50µm.

Supplementary Figure 4. (A) FACS plot showing the results of sorting from total adult kidney to generate fractions for MSC culture. Lin^{-}CD31^{-} cells were sorted into three fractions based on expression of CD24 and Sca-1 and collected for further analysis. MSC-like cultures could only be generated from the CD31^{-}Lin^{-}CD24^{lo}Sca1^{+} fraction. (B) FACS plot showing freshly isolated EGFP^{+} (Hoxb7^{+}) cells from the Hoxb7/EGFP mice has low level of CD24 expression (CD24^{lo}).

Supplementary Figure 5. Graph showing the comparison of population doubling time (PDTs) between Hoxb7-derived and bulk cultured kidney MSC-like cells at passage 3, 5, 9 and 13. At passage 3 the average population doubling time is 3.1 days for Hoxb7-derived MSCs and 5.0 days for bulk cultured MSCs.

Supplementary Figure 6. Lack of F4/80 expression in the Hoxb7GFP^{lo} fraction. (A) Reanalysis of the total kidney F4/80^{+} fraction for GFP vs EpCAM reveals no overlap between the GFP^{+} population and the macrophage population. (B) Immunofluroescence for Hoxb7GFP (green), F4/80 (red) and DAPI (blue) confirms no spatial overlap between F4/80^{+}

cells and either the GFP^{+} or GFP^{lo} (arrowhead) population of the PND5 kidney. Scale bar: 20μm.

Supplementary Figure 7. *In situ* hybridization of TS28 (PND6) mouse kidney (GUDMAP:14084; www.gudmap.org) show *Wnt4* expression (arrowhead) is seen in the interstitial compartment of the medulla but also in occasional collecting duct epithelial cells in the tip of the papilla at PND6. Scale bar: 100μm.

Supplementary Figure 8. (A) Low resolution confocal images show Wnt4 (EGFP) and TdTomato (red) distribution in cortex, medulla, papilla base and tip of the adult kidney (PND49). Scale bar: 20 μm. (B) Semi-quantification data show the number of $TdTomato^{+}$ (and $EGFP^{+}$) cells either within $Aqp2^{+}$ (collecting duct epithelial) or $Apq2^{-}$ population, confirming no $EGFP^{+}$ ($TdTomato^{+}$) cells are detected within the collecting duct epithelial at E19.5 but at PND49 8% of $Aqp2^{+}$ cells are $TdTomato^{+}$ ($EGFP^{+}$).

Supplementary Figure 9. Evidence for convergent extension in the elongation of the neonatal collecting duct. Immunofluorescence for GFP (green) and DAPI (blue), showing $Hoxb7GFP^{+}$ tubular structures in PND0 and PND15 mouse kidney. Dashed line outlines the edges of each tubule. Scale bar: 20μm.

Supplementary Table 1. List of primers used for qPCR.