KIDNEY CANCER

Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors

Matthew D. Young[1]*, Thomas J. Mitchell[1,2,3]*, Felipe A. Vieira Braga[1]*, Maxine G. B. Tran[4,5], Benjamin J. Stewart[6], John R. Ferdinand[6], Grace Collord[1,2,7], Rachel A. Botting[8], Dorin-Mirel Popescu[8], Kevin W. Loudon[6], Roser Vento-Tormo[1], Emily Stephenson[8], Alex Cagan[1], Sarah J. Farndon[1,9,10], Martin Del Castillo Velasco-Herrera[1], Charlotte Guzzo[1], Nathan Richoz[6], Lira Mamanova[1], Tevita Aho[2], James N. Armitage[3], Antony C. P. Riddick[3], Imran Mushtaq[9], Stephen Farrell[2], Dyanne Rampling[9], James Nicholson[2,7], Andrew Filby[8], Johanna Burge[2], Steven Lisgo[11], Patrick H. Maxwell[12], Susan Lindsay[11], Anne Y. Warren[2], Grant D. Stewart[2,3], Neil Sebire[9,10], Nicholas Coleman[2,13], Muzlifah Haniffa[8,14]†, Sarah A. Teichmann[1]†, Menna Clatworthy[2,6]†, Sam Behjati[1,2,7]†

Messenger RNA encodes cellular function and phenotype. In the context of human cancer, it defines the identities of malignant cells and the diversity of tumor tissue. We studied 72,501 single-cell transcriptomes of human renal tumors and normal tissue from fetal, pediatric, and adult kidneys. We matched childhood Wilms tumor with specific fetal cell types, thus providing evidence for the hypothesis that Wilms tumor cells are aberrant fetal cells. In adult renal cell carcinoma, we identified a canonical cancer transcriptome that matched a little-known subtype of proximal convoluted tubular cell. Analyses of the tumor composition defined cancer-associated normal cells and delineated a complex vascular endothelial growth factor (VEGF) signaling circuit. Our findings reveal the precise cellular identities and compositions of human kidney tumors.

Cancer cell identity is defined by morphological appearance, tissue context, and marker gene expression. Single-cell transcriptomics refines this cellular identity on the basis of a comprehensive and quantitative readout of mRNA. Precise cellular transcriptomes may reveal a tumor's cell of origin and the transcriptional trajectories underpinning malignant transformation (*1*).

We sought to define the identities of normal and cancerous human kidney cells from a catalog of 72,501 single kidney cell transcriptomes, integrated with tumor whole-genome DNA sequences (*2*). We studied Wilms tumor (n = 3 specimens), clear cell renal cell carcinoma (ccRCC) (n = 3), and papillary renal cell carcinoma (pRCC) (n = 1) in relation to healthy fetal (n = 2), pediatric (n = 3), adolescent (n = 2), and adult (n = 5) kidneys, as well as to ureters (n = 4) (table S1).

Normal tissue biopsies were taken from macroscopically normal regions of kidneys resected because of cancer (n = 10 samples) or for transplantation (n = 2 samples). We performed technical replicates of each biopsy and prepared biological replicates, where clinically permissible (table S1). We processed kidneys immediately after resection, generating single-cell solutions enriched for viable cells. We derived counts of mRNA molecules in each cell for further analyses, subject to quality control (*2*).

We split 72,501 fetal, normal, and tumor cells into immune and nonimmune cell compartments (fig. S1). Using a community detection algorithm (*2*), we further segregated transcriptomes into distinct clusters of cells (table S2). We next generated a reference map of normal mature and fetal cells, assigning an identity to each cluster, by cross-referencing cluster-defining transcripts with canonical markers curated from the literature (table S3). Ambiguous clusters were not included in the reference map and are presented in figs. S2 to S8. Highly specific cluster-defining transcripts (potential cell markers) are appended (table S4).

Among 42,809 nonmalignant cells, 37,951 mature kidney cells represented epithelial cells from distinct micro-anatomical regions of the nephron, with a large proportion of proximal tubular cells (Fig. 1, A to C, and fig. S4). Furthermore, there were fibroblasts, myofibroblasts, and vascular endothelial cells (glomerular endothelium and ascending and descending vasa recta) (Fig. 1D and fig. S2). Fetal cells (4858) grouped into developing nephron cells [ureteric bud (UB), cap mesenchyme (CM), and primitive vesicle (PV) cells] and fibroblasts, myofibroblasts, vascular endothelial cells, and ganglion cells (Fig. 2, A to C, and fig. S5).

To determine transcriptional programs underlying nephrogenesis, we identified transcription factors differentially expressed in UB cells versus CM and PV cells (Fig. 2D). Furthermore, we applied pseudotiming methods to identify transcription factors that define the transition from CM to PV (Fig. 2D). Together, these analyses identified both established and previously unknown transcription factors associated with nephron development, included as a reference for subsequent analyses of malignancy (table S5).

Having established the single-cell landscape of healthy kidneys, we characterized the cellular identities of 6333 nonimmune (fig. S7) and 17,821 immune (fig. S8) tumor cells from Wilms tumor (n = 3), ccRCC (n = 3), and pRCC (n = 1) (table S1). Children had received neoadjuvant cytotoxic treatment before nephrectomy, per British practice. Although this pretreatment reduced yield (table S6), recovered cells represent therapeutically relevant surviving cancer cells that determine the degree of adjuvant cytotoxic chemotherapy required (*3*). We used logistic regression to quantify the similarity between tumor and normal cell clusters, validated through intrinsic control populations (*2*). That is, the model found that myofibroblasts from tumors matched myofibroblasts from mature and fetal kidneys (Fig. 3A) and found no match for mast cells, a negative control population inserted into the training data.

This similarity metric may be obfuscated by the phenotypic plasticity of tumor cells. We therefore developed a method to genotype individual cancer cells from mRNA reads by using somatic copy number changes (table S7 and fig. S9) defined by whole-genome sequencing (fig. S10). We validated genotyping calls by phasing single-nucleotide polymorphisms across segments with altered copy numbers, testing for the presence of somatic single-nucleotide variants, and comparing with control populations (figs. S11 to S14).

Integrating genotyping and similarity analyses, we found that Wilms cells resembled normal fetal cells, showing that Wilms cells represent aberrant fetal cells. We found different populations of Wilms tumor that matched UB and PV cells (specific developing nephron populations) (Fig. 3A). One cluster (designated WF), composed of Wilms cancer cells and noncancerous ccRCC fibroblasts, exhibited a fibroblast-myofibroblast transcriptome. In one case, we obtained an anatomically

[1]Wellcome Sanger Institute, Hinxton CB10 1SA, UK. [2]Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 0QQ, UK. [3]Department of Surgery, University of Cambridge, Cambridge CB2 0QQ, UK. [4]UCL Division of Surgery and Interventional Science, Royal Free Hospital, London NW3 2PS, UK. [5]Specialist Centre for Kidney Cancer, Royal Free Hospital, London NW3 2PS, UK. [6]Molecular Immunity Unit, Department of Medicine, University of Cambridge, MRC Laboratory of Molecular Biology, Cambridge CB2 0QQ, UK. [7]Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK. [8]Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. [9]Great Ormond Street Hospital for Children NHS Foundation Trust, London WC1N 3JH, UK. [10]UCL Great Ormond Street Hospital Institute of Child Health, London WC1N 1E, UK. [11]Human Developmental Biology Resource, Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne NE1 3BZ, UK. [12]Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, UK. [13]Department of Pathology, University of Cambridge, Cambridge CB2 1QP, UK. [14]Department of Dermatology, Royal Victoria Infirmary, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne NE1 4LP, UK.
*These authors contributed equally to this work.
†Corresponding author. Email: sb31@sanger.ac.uk (S.B.); mrc38@medschl.cam.ac.uk (M.C.); st9@sanger.ac.uk (S.A.T.); m.a.haniffa@newcastle.ac.uk (M.H.)

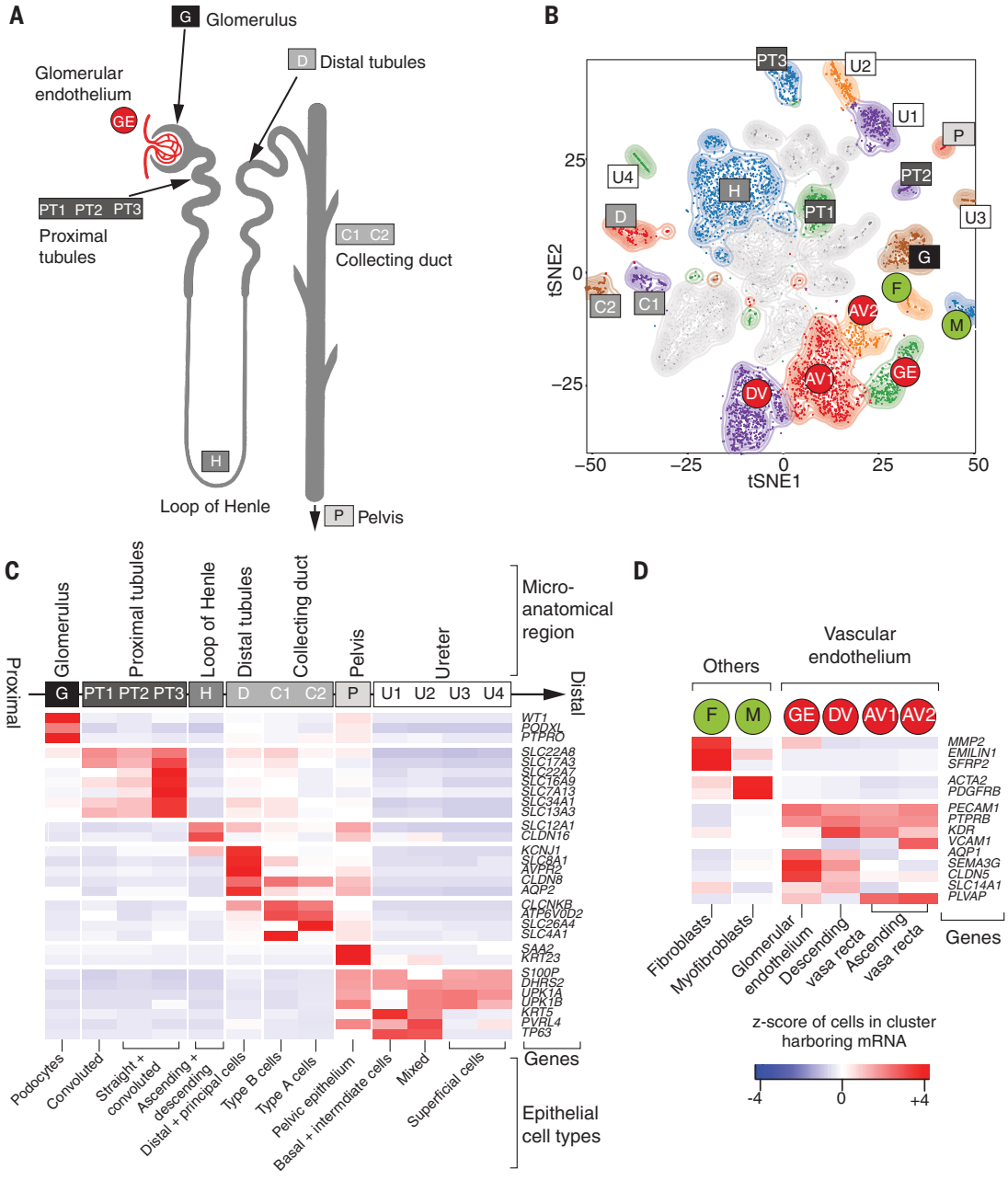

Fig. 1. Canonical cell types in normal human kidneys. (**A**) Illustration of nephron anatomy with cell clusters marked. (**B**) t-SNE (*t*-distributed stochastic neighbor embedding) representation of 8,707 normal epithelial and vascular cells. Clusters are colored, distinctively labeled, and emphasized with density contours. Ambiguous clusters are de-emphasized and fully shown in fig. S2. (**C**) Expression of canonical nephron-specific genes (table S3) in clusters from (A). Colors give the fraction of cells expressing each gene in a cluster, scaled to have a mean of 0 and SD of 1 across all clusters. (**D**) Expression of clusters in (A) not shown in (C) and their canonical genes.

separate perilobar nephrogenic rest, thought to represent a precursor lesion of Wilms tumor. We observed that, like Wilms cancer cells, the nephrogenic rest resembled UB or PV cells. This finding suggests that the potential to generate the different cell states of the fetal nephron is acquired early or was not lost by the developing Wilms cancer, although the basis for this conclusion is only one sample.

To validate the cellular identity of Wilms cells, we interrogated bulk transcriptomes of an independent series of 124 Wilms tumors for cellular signatures of UB and PV (*4*, *5*). We extracted specific markers expressed within UB or PV cells and unexpressed within nontumor cells (table S8) (*2*) and probed bulk transcriptomes for these cluster-defining transcripts. As comparators to Wilms tumor transcriptomes, we included fetal, pediatric, and adult normal tissue bulk transcriptomes (n = 135) and other childhood kidney tumors: 17 congenital mesoblastic nephroma and 65 malignant rhabdoid tumors. Corroborating the presence of these cells in Wilms tumor, signatures of PV and UB cells were seen in, and confined to, Wilms tumor and normal fetal tissues (Fig. 3B).

Placing Wilms tumor cells in pseudotime revealed two transcriptional programs emanating from the UB: one branch describing the development predominantly of nephrogenic rest cells and the other of Wilms cancer cells (Fig. 3C). The transcription factors underpinning these two programs (Fig. 3D and table S9) and normal nephrogenesis overlapped significantly ($P < 10^{-4}$; hypergeometric test). This indicates that developmental relationships exist among Wilms tumor cells that have been adopted from normal nephrogenesis. Our analyses reveal the plasticity and fetal identity of Wilms cells and transcriptionally define developmental cell states and trajectories that may harbor targetable vulnerabilities.

Next we studied ccRCC and pRCC (type 1), including one case of von Hippel–Lindau disease–related ccRCC (table S1). Matching ccRCC and

pRCC with normal mature cells, we found that they retained transcriptional features of cluster PT1, a specific subtype of convoluted proximal tubular cell (Fig. 4A). Most (six of seven) ccRCC clusters and all pRCC cells matched this particular PT1 cell, indicating that it represents an RCC cell state that transcends the diversity of RCC cells within and across tumors. Little is known about the nearest normal cell correlate of RCC, the PT1 cell, which has been found to become more abundant in inflamed renal tissue (*6*).

To validate the identity of the PT1 signature in RCC, we made use of the observation that PT1 cells were defined by SLC17A3 and VCAM1 with the absence of SLC7A13 within our data (Fig. 4B and fig. S2). We measured these transcripts in an independent series of 1019 publicly available bulk kidney tumor and normal tissue transcriptomes. High expression of SLC17A3 mRNA distinguished ccRCC and pRCC (types 1 and 2) from other types of RCC ($P < 10^{-4}$; Mann-Whitney test), whereas SLC7A13 mRNA was significantly depleted in ccRCC and pRCC bulk transcriptomes versus normal transcriptomes ($P < 10^{-4}$; Mann-Whitney test), as were mRNAs representing other regions of the nephron (Fig. 4B). VCAM1 expression, specific to PT1 within proximal tubules, was also significantly elevated across RCC bulk transcriptomes ($P < 10^{-4}$; Mann-Whitney test) (Fig. 4B), with each individual RCC tumor exhibiting PT1 features (fig. S15). Confocal microscopy demonstrated colocalization of VCAM1 and SLC17A3 in CA9^{+} cells, CA9 being

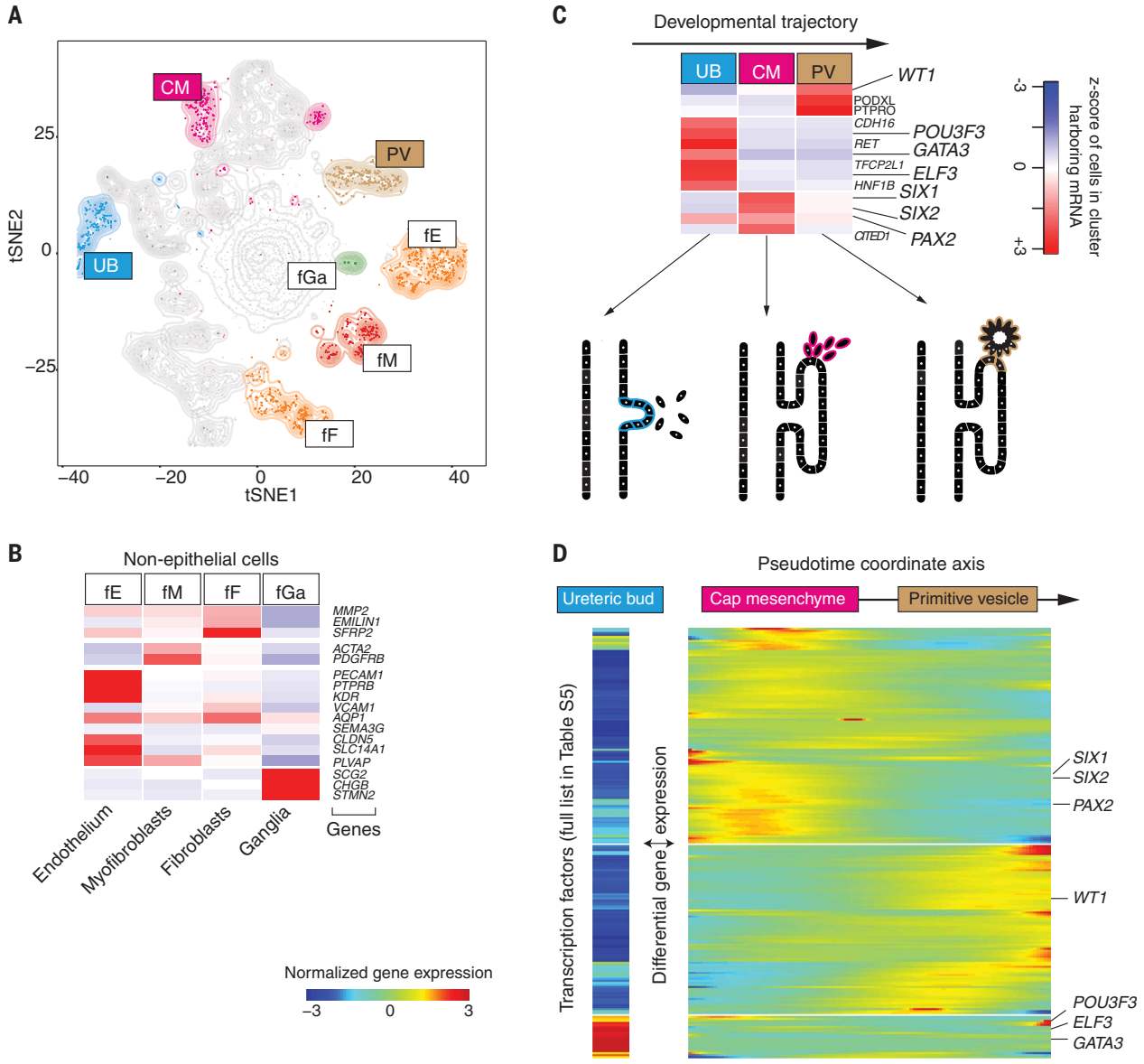

Fig. 2. Fetal cell types and nephrogenesis. (**A**) t-SNE representation of 4,858 fetal epithelial and vascular cells, colored and labeled as in Fig. 1B. fE, fetal endothelium; fGa, fetal ganglia; fM, fetal myofibroblasts; fF, fetal fibroblasts. (**B**) Expression of markers of clusters in (A), colored as in Fig. 1C. (**C**) Expression of nephrogenesis markers from clusters in (A) with illustration of nephron development. Formation of nephrons emanates from the UB, which induces condensation of the overlying mesenchyme into the CM. The CM then forms the PV, the precursor of the glomerulus. The tubular system grows out from both ends of the fetal nephron: the UB and the PV. (**D**) The expression of transcription factors that vary significantly ($P < 0.01$; likelihood ratio test) along the pseudotime trajectory defined by using the CM and PV cells from (C) or that are differentially expressed in UB versus CM and PV cells. UB expression is shown in a separate block on the left. Within the right block, pseudotime increases from left to right and rows are clustered and grouped by hierarchical clustering, with canonical transcription factors of nephrogenesis highlighted (see table S6).

a specific marker of ccRCC cells (Fig. 4C). Furthermore, we studied the earliest precursor lesions of ccRCC: CA9$^+$ proximal tubular cells residing in morphologically normal kidney tissue, predisposed to ccRCC through pathogenic germline mutation of *VHL*. Examining tissue from three individuals, we identified CA9$^+$ VCAM1$^+$ clusters of proximal tubular cells (Fig. 4D). Similarly, tumors arising in these kidneys harbored CA9$^+$ VCAM1$^+$ cells (fig. S16). As expected, VCAM1 was otherwise sparsely expressed on proximal tubular cells. Together, these observations substantiate our proposition that PT1 cells are the nearest normal cell correlate of ccRCC cells. The presence of the PT1 signature in both ccRCC and pRCC may indicate a common origin of these tumors with divergent fates.

Apart from the PT1 signature in pRCC and ccRCC, we found that one ccRCC cell cluster (cR7) matched PT3 cells and that pRCC cells exhibited an additional, weaker match with collecting duct cells (Fig. 4A). Neither signal was enriched in bulk transcriptomes (Fig. 4B). As our study was confined to type 1 pRCC, it is possible that we missed other pRCC cell types.

Finally, we dissected the tumor microenvironment occupied by cancer-associated normal cells, comprised of immune cells, fibroblasts, myofibroblasts, and vascular endothelial cells (predominately ascending vasa recta) (figs. S7, S8, and S17). Within these groups, we studied vascular endothelial growth factor (VEGF) signaling, an established target in RCC treatment (*7*, *8*). The VEGF signaling circuit in renal tumors involves VEGFA secretion from RCC cells, resulting in a response from endothelial cells (*7*, *8*). Measuring expression of the key components of VEGF signaling, we identified tumor-infiltrating macrophages as a further source of VEGFA (fig. S18A), as confirmed by confocal microscopy of ccRCC cells and flow cytometry of an independent ccRCC tumor (fig. S18, B to D). VEGF signaling receptors (KDR, FLT1, and FLT4) were expressed mainly by one population of ascending vasa recta cells (fig. S18A, cluster tE1). The other ascending vasa recta cluster, tE2 (fig. S18A), exhibited lymphangiogenic VEGFC and FLT1. Furthermore, tE2 endothelial cells expressed high levels of ACKR1, a marker of venular endothelium promoting tissue migration of immune cells (*9*). Overall, these findings delina complex VEGF signaling circuit within RCC tissue.

By identifying specific normal cell correlates of renal cancer cells, this study moves our understanding of these malignancies beyond a notion of "fetalness" or an approximate micro-anatomical

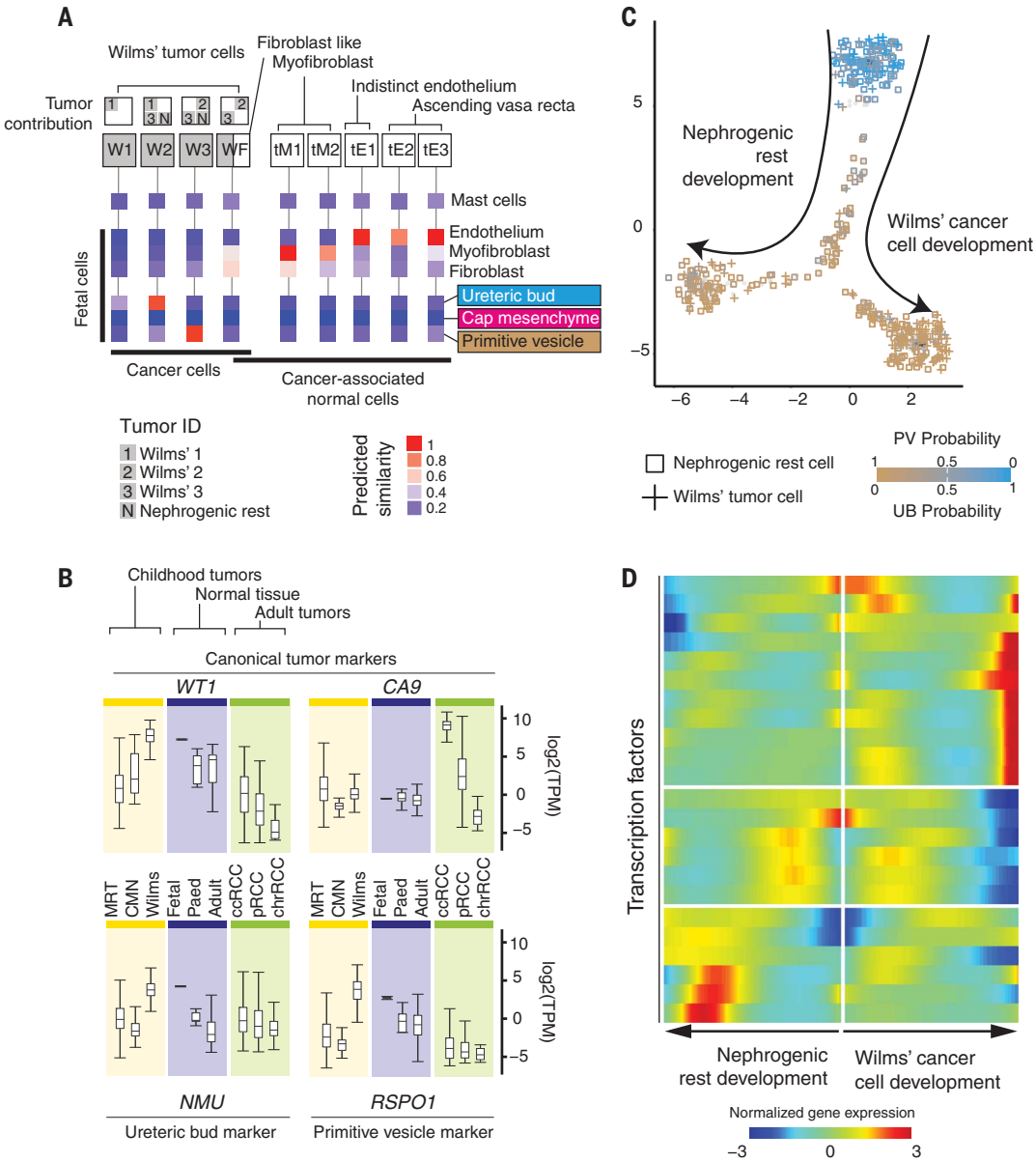

Fig. 3. Matching childhood tumors with normal fetal cells. (**A**) Similarity of Wilms tumor and cancer-associated normal cells to the reference fetal kidney map (Fig. 2A), with mast cells added as a negative control. Square boxes indicate sample contribution. Colors represent the probability that the cluster identified in the column header is "similar" to the fetal cluster identified by the row label (*2*). tM1 and tM2, tumor myofibroblast clusters 1 and 2; tE1 to tE3, tumor endothelial clusters 1 to 3. (**B**) Expression of canonical tumor markers and representative UB- and PV-specific genes (table S8) in bulk transcriptomes of childhood cancers (yellow), normal tissue (blue), or adult cancers (green). MRT, malignant rhabdoid tumor; CMN, congenital mesoblastic nephroma. As positive controls, canonical tumor markers are shown: WT1, Wilms tumor, and CA9, ccRCC. TPM, transcripts per kilobase million. (**C**) Pseudotime trajectory of all Wilms tumor and nephrogenic rest cells. Color indicates the similarity of each cell to the PV or UB fetal population. Jitter has been added to each point's position, with the original position plotted underneath in black (*2*). (**D**) Transcription factors identified as varying significantly along the pseudotime trajectory in (C). The center of the heat map corresponds to the cells at the top of (C), and the map then proceeds left and right along the arrows shown in (C).

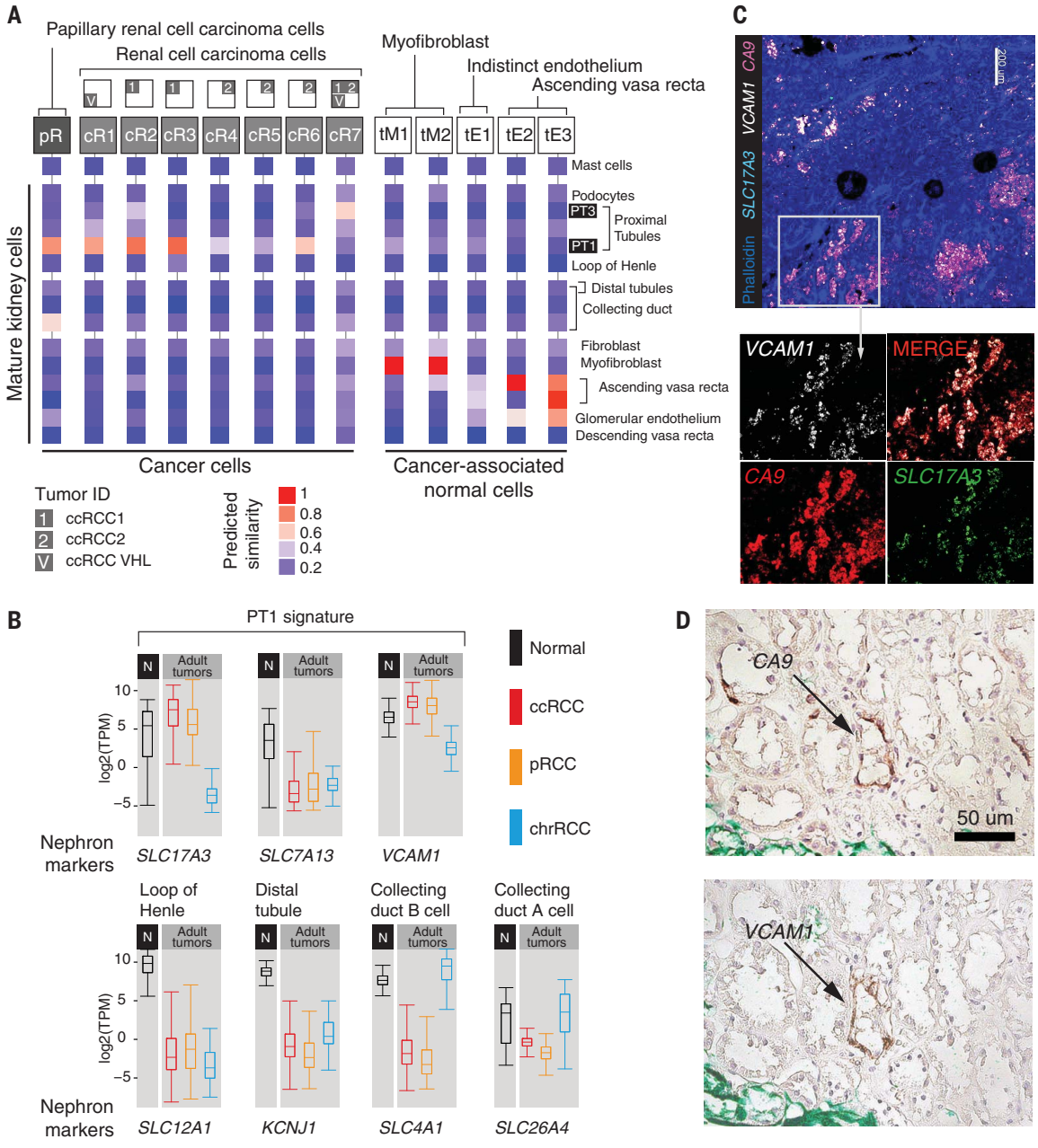

Fig. 4. Matching adult tumors with normal mature kidney cells. (**A**) Similarity of adult cancer and cancer-associated normal cells to the mature kidney reference map (Fig. 1B), with mast cells added as a negative control. Square boxes indicate sample contribution. Colors represent the probability that the cluster identified in the column header is "similar" to the normal cluster identified by the row label (*2*). pR, pRCC cell cluster; cR1 to cR7, ccRRC cell clusters. (**B**) Expression of nephron-specific genes in bulk transcriptomes as in Fig. 3B. pRCC samples are both types 1 and 2. (**C**) Confocal microscopy showing colocalization of PT1 markers (VCAM1 and SLC17A3) in ccRCC cells (CA9). (**D**) Staining of a proximal tubular ccRCC precursor lesion (CA9) for the PT1 marker VCAM1.

region to a precise cellular, molecularly quantitative resolution. Our findings portray the peak incidence of Wilms tumor in early childhood as a corruption of fetal nephrogenesis, in contrast to the lifelong development of RCC in mature kidneys. Our study provides a scalable experimental strategy for determining the identity of human cancer cells.

REFERENCES AND NOTES

1. C. Ziegenhain *et al.*, *Mol. Cell* **65**, 631–643.e4 (2017).
2. Materials and methods are available as supplementary materials.
3. K. Pritchard-Jones *et al.*, *Lancet* **386**, 1156–1164 (2015).
4. E. J. Perlman *et al.*, *Nat. Commun.* **6**, 10013 (2015).
5. S. Gadd *et al.*, *Nat. Genet.* **49**, 1487–1494 (2017).
6. D. Seron, J. S. Cameron, D. O. Haskard, *Nephrol. Dial. Transplant.* **6**, 917–922 (1991).
7. B. Ljungberg *et al.*, *Eur. Urol.* **67**, 913–924 (2015).
8. S. Fernández-Pello *et al.*, *Eur. Urol.* **71**, 426–436 (2017).
9. A. Thiriot *et al.*, *BMC Biol.* **15**, 45 (2017).

ACKNOWLEDGMENTS

We thank M. Stratton, P. Campbell, D. Rowitch, M. Gessler, and M. Ramakrishna for review of the manuscript and M. Gerstung and V. Svensson for advice regarding logistic regression. We are indebted to our patients and their families for participating in this research. **Funding:** This experiment was principally funded by the St. Baldrick's Foundation (Robert J. Arceci International Award to S.B.). Additional funding was received from Wellcome (S.B., M.H., G.C., C.G., P.H.M.), Cambridge Biomedical Research Campus (biobanking infrastructure; M.C.), CRUK Cambridge Centre (biobanking infrastructure), the NIHR Blood and Transplant Research Unit (M.C.), MRC (M.C.), Arthritis Research UK (M.C.), The Lister Institute for Preventative Medicine (M.H.), NIHR and Newcastle Biomedical Research Centre (M.H.), the ISAC SRL-EL program (A.F.), joint Wellcome Trust–MRC funding (S.Lis., S.Lin.), Kidney Cancer UK (M.G.B.T.), Facing up 2 Kidney Cancer (M.G.B.T.), EMBO (R.V.-T.), the Human Frontier Science Program (R.V.-T.), and Children with Cancer UK (S.J.F.). **Author contributions:** S.B. conceived of the experiment. M.D.Y., T.J.M., and S.B. analyzed the data, with contributions from F.A.V.B., B.J.S., M.D.C.V.-H., G.C., and M.C. Samples were curated and/or experiments were performed by F.A.V.B., J.R.F., M.G.B.T., P.H.M., R.A.B., D.-M.P., R.V.-T., E.S., K.W.L., S.J.F., C.G., N.R., L.M., T.A., J.N.A., A.C.P.R., I.M., S.F., D.R., J.N., A.F., J.B., S.Lis., S.Lin., and G.D.S. Pathological expertise was provided by A.Y.W., N.S., and N.C. A.C. created illustrations. T.J.M., M.D.Y., and S.B. wrote the manuscript. M.H., S.A.T., M.C., and S.B. codirected the study. **Competing interests:** P.H.M. discloses paid consultancy work for Mission Therapeutics and stock ownership in ReOx plc. A.Y.W. attended a meeting of the Roche UK Scientific Advisory

Board on PDL-1 testing in renal cell carcinoma. G.D.S. discloses paid consultancy work for Pfizer, EUSA Pharma, and Cambridge Medical and an advisory board role for the latter. **Data and materials availability:** Raw sequencing data have been deposited in the European Genome-phenome Archive (EGA) under study IDs EGAS00001002171, EGAS00001002486, EGAS00001002325, and EGAS00001002553. Sample-specific identifiers can be found in tables S6 and S10, a table of mapped unique molecular identifier counts for each cell and gene combination in data S1, and metadata about each cell in table S11. The code necessary to perform the analysis and generate figures can be obtained from https://github.com/constantAmateur/scKidneyTumors.

SUPPLEMENTARY MATERIALS
www.sciencemag.org/content/361/6402/594/suppl/DC1

Materials and Methods
Figs. S1 to S19
Tables S1 to S12
References (*10*–*46*)
Data S1

1 February 2018; accepted 2 July 2018
10.1126/science.aat1699

Science

Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors

Matthew D. Young, Thomas J. Mitchell, Felipe A. Vieira Braga, Maxine G. B. Tran, Benjamin J. Stewart, John R. Ferdinand, Grace Collord, Rachel A. Botting, Dorin-Mirel Popescu, Kevin W. Loudon, Roser Vento-Tormo, Emily Stephenson, Alex Cagan, Sarah J. Farndon, Martin Del Castillo Velasco-Herrera, Charlotte Guzzo, Nathan Richoz, Lira Mamanova, Tevita Aho, James N. Armitage, Antony C. P. Riddick, Imran Mushtaq, Stephen Farrell, Dyanne Rampling, James Nicholson, Andrew Filby, Johanna Burge, Steven Lisgo, Patrick H. Maxwell, Susan Lindsay, Anne Y. Warren, Grant D. Stewart, Neil Sebire, Nicholas Coleman, Muzlifah Haniffa, Sarah A. Teichmann, Menna Clatworthy and Sam Behjati

Science **361** (6402), 594-599.
DOI: 10.1126/science.aat1699

Pediatric and adult kidney tumors differ

Understanding tumor origins and the similarities and differences between organ-specific cancers is important for determining treatment options. Young *et al.* generated more than 72,000 single-cell transcriptomes from healthy and cancerous human kidneys. From these data, they determined that Wilms tumor, a pediatric kidney cancer, originates from aberrant fetal cells, whereas adult kidney cancers are likely derived from a specific subtype of proximal convoluted tubular cell.

Science, this issue p. 594