Cellular mRNA signals in human kidney tumors

Matthew D Young[†*1], Thomas J Mitchell[†1,2,3], Lars MC Custers[†4], Eleonora Khabirova[1], Thomas RW Oliver[1,2,5], Federico Comitani[6,7,8], Alice Piapi[9], Eva Bugallo Blanco[9], Christine Thevanesan[9], Kenny Roberts[1], Felipe A Vieira Braga[10], Tim HH Coorens[1], Ignacio Del Valle[9], Anna Wilbrey-Clark[1], Lira Mamanova[1], Grant D Stewart[2,3], Vincent J Gnagapragasam[2,3,12], Dyanne Rampling[11], Neil Sebire[9,11], Nicholas Coleman[2,5], Liz Hook[2,5], Anne Warren[2], Muzlifah Haniffa[1,13,14], Marcel Kool[15-17], Stefan M Pfister[15-17], John C Achermann[9], Xiaoling He[18,19], Roger A Barker[18,19], Adam Shlien[6,7,8], Omer A Bayraktar[1], Sarah Teichmann[1], Kerstin B Meyer[1], Jarno Drost[4*], Karin Straathof[9,11*], Sam Behjati[1,2,20*]

Affiliations

[1]Wellcome Sanger Institute, Hinxton CB10 1SA, UK.

[2]Cambridge University Hospitals NHS Foundation Trust, Cambridge, CB2 0QQ, UK.

[3]Academic Urology Group, Department of Surgery, University of Cambridge, Cambridge, CB2 0QQ, UK.

[4]Oncode Institute, Princess Máxima Center for Pediatric Oncology, 3584 CS Utrecht, The Netherlands. [5]Department of Pathology, University of Cambridge, Cambridge, CB2 1QP, UK.

[6]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada.

[7]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada.

[8]Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada.

[9]UCL Great Ormond Street Hospital Institute of Child Health, London WC1N 1EH, UK.

[10]Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands.

[11]Great Ormond Street Hospital for Children NHS Foundation Trust, London WC1N 3JH, UK.

[12]Cambridge Urology Translational Research and Clinical Trials office, Cambridge Biomedical Campus Cambridge CB2 0QQ University of Cambridge

[13]Department of Dermatology and NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, NE1 4LP, UK.

[13]Intitute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, NE1 4HH, UK.

[15]Hopp Children´s Cancer Center Heidelberg (KiTZ), 69120 Heidelberg Germany.

[16]German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Division of

Pediatric Neurooncology, 69120 Heidelberg, Germany.

[17]Heidelberg University Hospital, Department of Pediatric Hematology and Oncology, 69120 Heidelberg, Germany.

[18]MRC-WT Cambridge Stem Cell Institute, University of Cambridge, CB2 0QQ, Cambridge, UK.

[19]Department of Clinical Neuroscience, University of Cambridge, Cambridge, CB2 0QQ, UK.

[20]Department of Paediatrics, University of Cambridge, Cambridge, CB2 0QQ, UK.

[†]Joint first authors.

[*]Corresponding authors: sb31@sanger.ac.uk; my4@sanger.ac.uk; k.straathof@ucl.ac.uk; j.drost@prinsesmaximacentrum.nl

Abstract

Tumor cells may produce many of the same messenger RNAs (mRNA) as the cell they derive from. The relative abundance of these mRNAs, the transcriptiomic profile, may provide clues into the origin and development of tumors. Here we investigated the cellular origins of 1,300 childhood and adult renal tumors, spanning 7 different subtypes. We decomposed tumor bulk transcriptomes into single cell components, measuring the abundance of single cell derived reference "cellular signals" in each tumor. We quantified the extent to which each tumor utilized fetal cellular signals, finding that all childhood renal tumors are definitively fetal. This replaces the long-held presumption of "fetalness" with a precise, quantitative readout of immaturity. Analyzing cellular signals in each tumor type, we recapitulated previous findings for some, whilst providing novel insights into other, less well understood tumor types. For example, our analyses predicted fetal interstitial cells as the cell of origin of the infant kidney tumor, congenital mesoblastic nephroma, and demonstrate that another childhood kidney cancer, malignant rhabdoid tumor, arises from mesodermally derived cells in early development. We found remarkable uniformity in the cell signal of each tumor type, indicating the possible therapeutic and diagnostic utility of cellular signal decomposition. We demonstrated this utility with an example of a child with a cryptic renal tumor, which had not been identifiable by conventional diagnostic work-up but was clearly classified with our approach. Our findings provide a cellular definition of human renal tumors through an approach that is broadly applicable to human cancer.

Introduction

As cancer cells evolve from normal cells, they may retain messenger RNA (mRNA) characteristic of the cell of origin. In such cases, tumor transcriptomes contain information that describe the cell of origin itself, the tumor's differentiation state, or its pathway towards a cancer cell. Applied to human renal tumors, it is therefore conceivable that mRNAs reveal specific kidney cell types from which tumors arise, and delineate differentiation states of tumors, such as fetal developmental signals in childhood renal tumors.

Tumor bulk transcriptomes for most types of human cancer have been generated in the context of cancer genomics efforts of recent years, such as those conducted by the *International Cancer Genome Consortium* (ICGC) and the *Tumor Cancer Genome Atlas* (TCGA)[1,2]. A current major limitation of bulk tumor mRNA data is that transcripts cannot be assigned to their cellular source, which fundamentally limits the resolution of biological insight. For example, bulk mRNA analyses of clear cell renal cell carcinoma (ccRCC) broadly placed ccRCC cells on the spectrum of proximal tubular cells[3,4]. However, only when studied at single cell level, which enabled the direct quantitative comparison of single tumor and normal cell transcriptomes, could a specific subtype of proximal tubular cell be identified as the normal cell correlate of ccRCC cells[5]. Moreover, the masking of the cellular source of transcripts in bulk mRNA data blurs information about cancer-associated normal cells within the tumor microenvironment. As an example of this, VEGF signaling, an established therapeutic target in clear cell renal cell carcinoma (ccRCC), had been thought to originate from VEGF production by ccRCC cells[6]. However, recent analyses of the ccRCC microenvironment at cellular resolution revealed that mononuclear phagocytes were an additional source of VEGF-A in ccRCC[5].

The application of single cell mRNA sequencing in large patient cohorts is currently logistically challenging due to high resource requirements. An alternative approach to define cellular signals within tumor bulk with greater granularity is to identify the presence of single cell derived mRNA signals in bulk transcriptome, utilizing normal single cell

transcriptomes as a reference. Single cell reference data, generated by efforts collectively known as the *Human Cell Atlas*[7,8], have begun to provide quantitative transcriptional definitions of the normal cells that constitute the developing and mature human kidneys [5,9-13]. Using the entirety of these gene expression catalogues, rather than individual cell marker transcripts, we may therefore be able to identify single cell signals in bulk transcriptomes from kidney tumors.

Here, we studied normal single cell mRNA signals in bulk kidney tumor transcriptomes (n=1,258; **Figure 1a, Supplementary Table 1**). The central aim of our analyses was to derive normal cell correlates for each tumor type. In addition, we investigated developmental states of kidney tumors, the tumor micro-environment, and explored the diagnostic utility of single cell signals.

Results

Validity of cell signal assessment

To assess single cell mRNA signals in bulk tumor transcriptomes, we devised a method that fits raw bulk mRNA counts for the entire transcriptome – not just marker genes – to a weighted linear combination of transcriptomic signals derived from reference single cell data (**Figure 1a**). In order to derive the reference data, we combined single cell mRNA signals from previously published single cell mRNA data derived from human fetal, pediatric, and adult normal kidneys[5,13], as well as newly generated single cell mRNA reads from fetal kidney and adrenal glands (**Figure 1a, Supplementary Figures 1,2, Supplementary Table 2**).

A key requirement for our method was to reveal how well the reference single cell signals explain the bulk transcriptome. To this end we assigned unaccountable mRNAs to a flat transcriptomic signal (achieved via an intercept term, see **Methods**), which limits assignment of spurious signals when a bulk transcriptome differs from all signals in the reference. We termed this flat transcriptomic signal the “unexplained signal”, as it is a proxy for transcriptional patterns absent from the reference signals. We then fitted a range of

bulk transcriptomes using deliberately implausible sets of reference signals (**Supplementary Figure 3)**. This established the expected range of contributions of the unexplained signal in fitting bulk transcriptomes when an appropriate reference is absent.

The utility of this unexplained signal is illustrated by comparing our method to other approaches. We analyzed cell signals in bulk mRNA of purified normal human B-cells, human pre-B cell leukemia, and hematopoietic stem cells, using single cell transcriptomes from peripheral blood cells as a reference (**Figure 1b**). A method based on assigning cell signals using markers of reference cell populations[14] was unable to differentiate between normal mature B cells and cancerous pre-B cell leukemia cells, giving both a mature B cell label (**Figure 1b**). Furthermore, the marker-based approach identified hematopoietic stem cells, for which no reference population was given, as other leucocytes (predominant signal dendritic cells).

A published deconvolution method, MuSiC[15], that uses not just markers, but entire transcriptomes from reference populations for cell label assignment, also failed to correctly identify the absence of a match for hematopoietic stem cells. Instead, it assigned a mixed B-cell and implausible renal proximal tubular cell identity to these hematopoietic stem cells. Our method, however, correctly identified the "unexplained signal" as the main signal in hematopoietic stem cells; that is, hematopoietic stem cells cannot be accounted for by the deliberately incomplete reference we provided. Similarly, MuSiC found an implausible renal cell signal in leukemia samples. By contrast, our method identified that pre-B cell leukemia samples resemble B-cells whilst more than ~35% (range 29-53%) of the leukemia bulk mRNA could not be accounted for by the incomplete reference. Taken together, these comparisons demonstrate that the unexplained signal metric of our method uniquely highlights when reference data is inappropriate, thus avoiding identification of false signals.

As an example of our method's performance, we assessed whether our method would distil the known cellular identity of cancer cells from bulk tissues, despite these being admixtures of cancer and normal cell populations. To this end, we assessed cell signals in 766 ccRCC transcriptomes from The Cancer Genome Atlas[16]. Our method could correctly identify the

signal of a specific proximal tubular cell population as the predominant cell signal in ccRCC bulk cancer transcriptomes (**Figure 1c, Supplementary Figure 4**).

Childhood tumors exhibit quantifiable fetal mRNA signals

For each tumor, we determined whether it exhibited a fetal or mature (i.e. post-natal) transcriptome, to guide the choice of reference in subsequent analyses. This was particularly relevant for childhood tumors which are believed to be of fetal origin. Evidence for this notion is largely lacking, with the exception of Wilms tumor, which has been likened to fetal nephron cells by cancer-to-normal cell mRNA comparison[5]

We calculated the immaturity by fitting each bulk transcriptome to a combined reference set composed of cellular signals from both mature and fetal kidney reference populations. The immaturity score was the sum of signals contributing to the bulk transcriptome derived from fetal kidney cells. Using this approach, we established a reference range of mature normal kidneys (**Figure 2a**). We demonstrated the validity of this range by scoring fetal kidney transcriptomes which lay significantly outside the mature range ($p=0.015$, Wilcoxon rank sum test).

We next calculated the same maturity score for individual tumors, which showed a clear signal of "fetalness" across all types of childhood kidney tumors (**Figure 2b,c**). This signal was mostly absent from adult tumors (**Figure 2b,c**), with the exception of a cohort of lethal chromophobe RCC that lay significantly outside the mature range ($p<10^{-4}$, Wilcoxon rank sum test)[16]. This prompted us to compare survival between adult tumors that lay within, or outwith, the normal maturity range (excluding the aforementioned lethal chromophobe tumors). This revealed that there was a significantly ($p=0.032$; log-rank test) reduced survival for patients with immature tumors (**Figure 2d**).

Congenital mesoblastic nephroma resembles fetal interstitial cells

CMN is a renal tumor of infants that has low metastatic potential. There are two morphological subtypes of CMN, classical and cellular variants[17]. Applying our method to define the contribution of different single cell signals in a cohort of CMN bulk transcriptomes (n=18), revealed a uniform signal of fibroblast-like, fetal interstitial cells across tumors (**Figure 3a, Supplementary Figure 5**), irrespective of morphological subtype. Of note, these fetal interstitial cells were characterized by expression of *NTRK3* and *EGFR* genes (**Figure 3b**), the principal oncogenes that drive CMN through activating structural variants[18]. Importantly, in a direct comparison of fetal interstitial cells *versus* mature renal fibroblasts within each sample, the interstitial cell signal was significantly higher ($p<10^{-4}$, one-sided Wilcoxon signed rank test), further corroborating the "fetalness" of CMN (**Figure 3c**). Furthermore, the interstitial cell signal was absent from, or not predominant in, control analyses indicating that it is does not simply represent artefacts or stromal signal (**Figure 3c).** To validate this interstitial cell signal in CMN, we subjected cells dissociated from a fresh CMN tumor specimen to single cell mRNA sequencing using the Chromium 10x platform. Tumor cell transcriptomes thus derived (**Figure 3d**) were compared to single cell clusters of normal fetal kidneys using previously developed quantitative approaches[5]. We found that CMN tumor cells matched the same interstitial cell population, validating the cell signal seen in bulk tumor tissue (**Figure 3e**).

Malignant rhabdoid tumors exhibit signals of mesodermal cells

Malignant rhabdoid tumor (MRT) is an aggressive, often fatal childhood cancer, that typically affects the kidney but may also occur at other sites. It is considered to be the extracranial counterpart of the CNS tumor, atypical teratoid/rhabdoid tumor (AT/RT). The principal, usually sole, driver event in MRT and AT/RT is biallelic inactivation of *SMARCB1*. Intracranial and extracranial rhabdoid tumors occasionally co-occur. In previous analyses of microRNA profiles, MRTs co-clustered with a range of tissues: neural crest derived tumors, cerebellum, and synovial sarcoma[19].

Assessing fetal renal single cell signals in 65 MRT bulk transcriptomes yielded an interstitial cell signal (**Figure 4a, Supplementary Figure 6**). However, when we studied MRT single cell transcriptomes, derived from an MRT expanded by a primary organoid culture (see **Methods**), MRT cell transcriptomes did not show any match within our fetal kidney reference (**Figure 4b**). This may indicate that the interstitial cell signal obtained in bulk represents a signal of the broad embryological lineage of the tumor, rather than a cell type. We therefore compared MRT cells against published reference cell populations of gastrulation embryos generated from mice[20], a developmental stage that is not accessible to study in humans. We found a specific match of MRT cells to paraxial mesoderm cells (**Figure 4c,d**). The validity of the murine cells as a reference was established by including control cells in the analysis; that is, renal endothelial cells matched, as expected, endothelial cells of the mouse embryo. Overall, therefore our analyses liken MRT transcriptomes to that of early embryonic mesodermal cells that do not have a direct counterpart in the fetal kidney.

We next investigated whether MRT, as previously suggested, additionally exhibited neural crest signals. Although we did not find murine neural crest signals in the MRT we studied at single cell resolution, it may be conceivable that a more extended survey of MRT would reveal subsets that do exhibit neural crest signals. To this end we re-examined MRT bulk transcriptomes for the presence of neural crest signals, by comparing these to reference signals derived human fetal adrenal glands which contain neural crest derived cells within the medulla. We performed single cell mRNA sequencing of two entire adrenal glands from a fetus of ~11 weeks' gestation to generate a census of adrenal cells, annotated using canonical markers described in the literature (**Supplementary Figure 1**). Again, MRT predominantly exhibited an interstitial cell signal (**Figure 4e, Supplementary Figure 7**), whereas control tumors, neuroblastoma, exhibited the expected medullary ganglion cell signal (**Figure 4e, Supplementary Figure 7**). Overall therefore our findings do not support the notion that MRTs exhibit neural signals. If indeed derived from the neural crest, it would indicate that MRTs form along the differentiation trajectory of neural crest to mesenchyme conversion, without leaving a transcriptional trace of their neural crest origin.

Fetal cell signals in Wilms tumor and clear cell sarcoma of the kidney

Wilms tumor is the most common childhood kidney cancer and is thought to arise from aberrant cells of the developing nephron. Assessing cell signals in a cohort of high risk, treatment-naïve Wilms cancers revealed a largely uniform fetal nephron signal derived from cap mesenchyme cells (**Figure 5a, Supplementary Figure 8**), a population of fetal renal cells that form precursors of glomeruli and the proximal tubular system. Interestingly, in a direct study of Wilms single cell transcriptomes derived from tumors pre-treated with cytotoxic chemotherapy, other developing nephron cell types, ureteric bud and primitive vesicle cells, predominated, consistent with selection by prior therapy.

Clear cell sarcoma of the kidney (CCSK) is a rare, at times aggressive childhood renal cancer that is treated as a high risk Wilms tumor in clinical practice [21]. Previous gene expression analyses of CCSK have suggested a neural phenotype for CCSK. Assessing normal cell signals in 13 CCSKs, we found that CCSK were dominated by an unexplained signal, indicating that the reference signals we provided did not account for the predominant cellular signal within CCSK bulk transcriptomes (**Figure 5b, Supplementary Figure 9**). We tested reference signals derived from fetal kidneys, mature kidneys, and fetal adrenal glands; the unexplained signal dominated in all cases. Interestingly, CCSK specific genes that had previously been interpreted as neural by gene ontology (GO) annotation, originated in fact from fetal renal cells (**Figure 5c**). Taken together, these findings neither confirm the neural cellular features of CCSK, nor identify a CCSK related cell signal within the fetal kidney. It is possible that CCSK represent transcriptionally grossly distorted renal cells, or that CCSK originates from non-renal cells.

Cell signals in, and composition of, adult kidney tumors

As discussed above, our analyses confirmed a previous finding that the predominant single cell signal in the most common types of adult renal cancer, clear cell RCC (ccRCC) and papillary RCC (pRCC), was derived from a specific subtype of proximal tubular cells, termed PT1 cell[5]. In addition, signal assessment also revealed some properties of the tumor microenvironment. We found a prominent vascular endothelial signal in ccRCCs (**Figure 6a,**

Supplementary Figure 10), but not in pRCCs. The downstream effects in RCC of inactivation of the von Hippel–Lindau gene and upregulation of vascular endothelial growth factors are well documented[22]. The prominent difference in the endothelial signal provides a read-out of this pathway, further explaining why anti-angiogenic treatments appear to be more effective in ccRCCs than in pRCCs[23].

Previous analyses of chromophobe cell renal cell carcinoma (ChRCC) have shown that ChRCC exhibit expression profiles of collecting duct cells [24]. Controversy exists as to whether the normal cell correlate of ChRCC is the type A or type B intercalated cells [3]. This is in part due to ChRCC retaining expression of both canonical markers of intercalated cells, SLC4A1 and SLC26A4 respectively (**Figure 6b**). Using our method of considering the entire transcriptomes of type A and type B cells, rather than just two markers, revealed a uniform type A signal across all chromophobe tumors (**Figure 6c, Supplementary Figure 11**), bar the lethal variant of so-called metabolically divergent tumors (**Figure 2b, Supplementary Figure 11**). The proliferation and active remodeling of type A cells has been demonstrated under conditions of systemic acidosis[25], lending further credence to their possible status as the cell of origin for ChRCCs.

Single cell signals provide diagnostic clues

An overarching finding of our study was that particular single cell signals were uniform in, and specific to, bulk transcriptomes from individual tumor types (**Figure 7a,b, Supplementary Figure 12**). Accordingly, cellular signal assessment of bulk transcriptomes may provide sensitive and specific diagnostic clues. To test this proposition, we examined cellular signals in the bulk transcriptome of a histologically undefinable metastatic primary renal tumor from an 11 year old boy. Following resection, the tumor was examined histologically locally and by international reference renal pathologists (**Figure 7c**). A definitive diagnosis could not be reached although an adult type renal cell carcinoma was favored. Nevertheless, the child was treated as a Wilms-like tumor, with cytotoxic chemotherapy and radiotherapy, following nephrectomy. He remains in complete remission one year following diagnosis, thus retrospectively suggesting a diagnosis of a Wilms-like tumor, as adult type carcinomas typically do not respond to such cytotoxic chemotherapy.

We performed bulk mRNA sequencing on tumor specimens from this patient. Assessment of mRNA signals in bulk tissue suggested that the tumor exhibited a fetal transcriptome with cellular signals consistent with a Wilms-like tumor (**Figure 7d,e**). The transcriptional diagnosis of a Wilms-like tumor was further substantiated by analyses of whole genome sequences. The tumor harbored classical somatic changes of Wilms, namely canonical *CTNNB1* and *KRAS* hotspot mutations and uniparental disomy of 11p (**Supplementary Figure 13**). By comparison, when we assessed single cell signals of an adult-type ccRCC that developed in a 15 year old adolescent, we found an overall mature transcriptome. Furthermore, the tumor exhibited the PT1 signal of ccRCC as well as a stark vascular endothelial typical of ccRCC (**Figure 7f,g**).

Discussion

We have determined normal cell signals in the major types of human renal tumors using a generalizable method that quantifies mRNA signals of single normal cells in bulk transcriptomes. In addition to providing specific insights into the normal cell correlates of individual tumor types, we have found that childhood tumors exhibit signals of human development. This has enabled us to replace the approximate notion of the “fetalness” of childhood renal tumors with quantitative transcriptional evidence that the entire spectrum of pediatric renal tumors represent an aberrant developmental state. As the compendium of single cell transcriptomes of human organs during fetal and post-natal life expands, our approach may reveal and define the developmental roots of childhood cancer more broadly.

A further finding of our study was that within each category, the majority of tumors exhibited remarkably uniform cellular signals. This indicates that there are overarching transcriptional features, beyond individual gene markers, that unite tumor entities despite underlying intra- and inter- tumor genetic heterogeneity. Therefore, cellular signals of renal tumors may lend themselves as diagnostic adjuncts, as illustrated here by our ability to resolve the identify of a histologically unclassifiable childhood tumor. Moreover, the cellular transcriptome itself may represent a therapeutic target that transcends individual patients,

if we had tools available to manipulate transcription in a predictable manner. This may be a particularly attractive approach for targeting transcriptional states of fetal cells retained in childhood cancer that are absent from normal post-natal tissues.

A central question that our findings raise is whether mRNA signals point to the cell of origin of tumors. It is conceivable that in some tumors malignant transformation entirely distorts and obliterates gene expression profiles of the cell of origin. For example, we found aberrant mRNA signals amongst high risk adult renal tumors that were characterized by an immature transcriptome and atypical cell type signals. These tumors may represent examples of what might be termed “dedifferentiated” tumors that have lost their cell of origin signal. Interestingly, our analysis suggests that in adult renal tumors the degree of transcriptional dedifferentiation correlates with overall survival. In other cases, it is more likely that the cell signals do directly capture the cell of origin. Our observation made a strong case in this regard for CMN originating from a specific population of fetal interstitial cells. The cellular signals of CMN and a direct comparison of CMN and normal single cell transcriptomes pointed to the same fetal cell type. Similarly, in adult tumors our findings suggest a common origin of chromophobe tumors in type A intercalated cells. Finally, in one tumor type, MRT, our findings are likely to point to the fetal lineage that MRTs represent, rather than a specific cell type, whilst demonstrating the absence of a neural crest identity previously assigned to these tumors.

Overall our findings attach specific cell labels to human renal tumors that are underpinned by quantitative molecular data obtained from single cell mRNA sequences, independent of the interpretation of marker genes. As reference data from single cell transcriptomes expand through efforts such as the *Human Cell Atlas*, it will be feasible to annotate existing large repositories of tumor bulk transcriptomes, to derive a cellular transcriptional definition of human cancer.

Acknowledgments

We acknowledge funding from Wellcome (Fellowships to S.B., K.S., J.C.A., core funding to the Sanger Institute, strategic award 211276/Z/18/Z). Additional support was provided by the St Baldrick's Foundation (S.B.), Great Ormond Street Hospital Children's Charity (J.C.A.), Great Ormond Street Hospital Biomedical Research Centre, Olivia Hodson Cancer Fund (K.S.), CRUK (Fellowship to T.J.M.), National Institute for Health Research (J.C.A., R.A.B.). The views expressed are those of the authors and not necessarily those of the National Health Service, National Institute for Health Research, or Department of Health. The CUH adult renal cancer sampling had infrastructure support from the Cancer Research UK Cambridge Cancer Centre Urological Malignancy Program and NIHR Biomedical Research Centre. J.D. acknowledges funding from the Dutch Cancer Society (KWF/Alpe d'HuZes Bas Mulder Award; KWF/Alpe d'HuZes, #10218), Foundation Children Cancer Free (KiKa #338, L.M.C.C.), Oncode Institute. We thank Dr Amos Burke for his contribution. We are grateful to our patients, young and old, for participating in our study.

Author Contributions

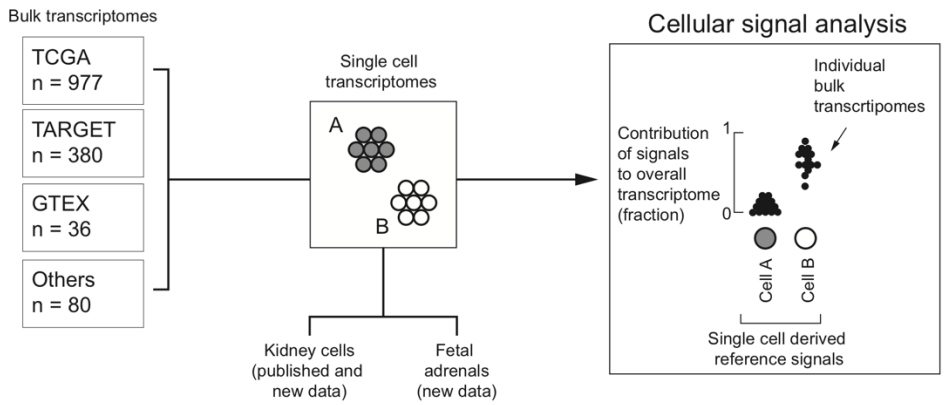
M.D.Y. and S.B. conceived of the experiment. M.D.Y. performed analyses, aided by T.J.M., E.K. and T.H.H.C. I.D.V. and J.C.A. provided expertise on adrenal gland analysis. L.M.C.C. performed organoid experiments with F.A.V.B. T.R.W.O., N.S., D.R., N.C., L.H., A.Y.W. provided pathological expertise. F.C. and A.S. provided clinical data. A.P., E.B.B., C.T., G.D.S., V.J.G., M.H., M.K., S.P., O.A.B., K.R. contributed to discussions and / or data. S.T. contributed fetal single cell data, together with K.M., R.A.B., X.H., A.W.C, L.M. S.B. directed the study, in conjunction with M.D.Y (method development), K.S. (single cell cancer work) and J.D (organoid work).

Declaration of Interests

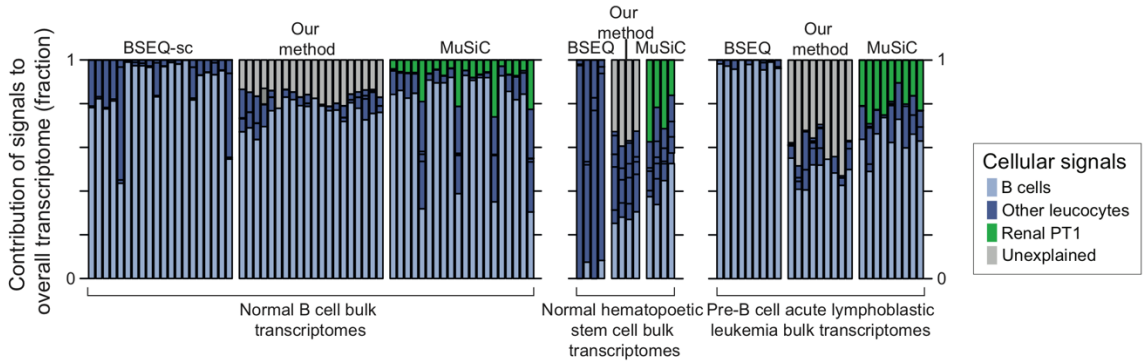
The authors declare no competing interests.

Figure 1

A


Bulk transcriptomes
TCGA
n = 977
TARGET
n = 380
GTEX
n = 36
Others
n = 80
Single cell
transcriptomes
A
B
Kidney cells
(published and
new data)
Fetal
adrenals
(new data)
Cellular signal analysis
Individual
bulk
transcrtipomes
Contribution
of signals
to overall
transcriptome
(fraction)
1
0
Cell A
Cell B
Single cell derived
reference signals

B


Contribution of signals to
overall transcriptome (fraction)
BSEQ-sc
Our
method
MuSiC
BSEQ
Our
method
MuSiC
BSEQ
Our
method
MuSiC
1
0
Normal B cell bulk
transcriptomes
Normal hematopoetic
stem cell bulk
transcriptomes
Pre-B cell acute lymphoblastic
leukemia bulk transcriptomes
Cellular signals
B cells
Other leucocytes
Renal PT1
Unexplained

C

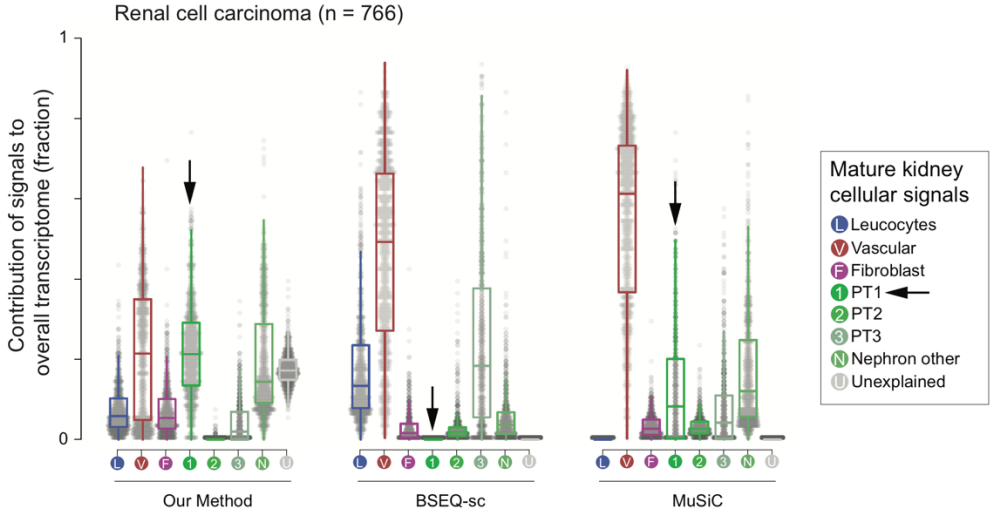

Renal cell carcinoma (n = 766)
Contribution of signals to
overall transcriptome (fraction)
1
0
L V F 1 2 3 N U
Our Method
BSEQ-sc
MuSiC
Mature kidney
cellular signals
L Leucocytes
V Vascular
F Fibroblast
1 PT1
2 PT2
3 PT3
N Nephron other
U Unexplained

Figure 1 – Methodology overview and validation

A. **Schematic of methodology:** Overview of the method to determine the contribution of single cell reference signals to explaining bulk RNA-seq transcriptomes. Bulk RNA-seq datasets (left, **Supplementary Table 1**) are combined with reference transcriptomic signals derived from single cell RNA-seq reference data (middle, **Supplementary Tables 2 and 3**) and the relative contribution of each reference signal in explaining the bulk RNA-seq transcriptomes is calculated (see Methods). The strength of the signals are normalized to a sum of 1, so each signal (corresponding to a population of single cells in a reference single cell RNA-seq dataset) is given a score between 0 and 1 for each bulk transcriptome, representing the relative strength of this signal in the bulk transcriptome in question.
B. **Method comparison on flow sorted immune-cells and ALL:** Comparison of two widely used "bulk deconvolution" methods (BSEQ-sc and MuSiC) to the new method developed here. Each bar represents a bulk transcriptome from a population of flow sorted adult B-cells (left), hematopoietic progenitors (middle), and pre-B acute lymphoblastic leukemia (right). The signal contribution is calculated for each sample using a reference signal set consisting of immune cells from adult peripheral blood mononuclear cells and proximal tubular cells from the mature (pediatric and adult) kidney (PT1 cells, included as a negative control). The relative contribution of each of these reference signals (plus an "unexplained signal" component where appropriate) is shown by the size and color of the stacked bars for each sample, as indicated by the legend on the far right. Raw signal contributions are listed in **Supplementary Table 4**.
C. **Ability of different methods to recover known RCC signal:** We applied our method, BSEQ-sc and MuSiC to a population of clear cell and papillary cell renal cell carcinomas, using reference signals derived from the mature normal kidney. This population of tumors is known to resemble the PT1 population of proximal tubular cells and have a strong vascular component. The relative contribution of each signal to each bulk RNA-seq sample is shown by the y-axis. The results for the three methods are split into blocks, with our method on the left, BSEQ-sc in the middle, and MuSiC on the right. Each signal type is labelled with an abbreviation and colored

as shown by the legend on the right. Signals are marked with a square for fetal kidney and circle for mature kidney. For our method, there is an additional signal which represents the unexplained signal for each sample (see Methods). Contributions due to vascular signals, fibroblast signals, leucocyte signals, and all non-proximal tubular nephron signals are aggregated together. Each signal/sample combination is represented by a single point and the distribution of relative signal contributions to the bulk transcriptomes are summarized with boxplots. Arrows mark the PT1 population which these tumors are known to resemble. Raw signal contributions are listed in **Supplementary Table 5**.

Figure 2

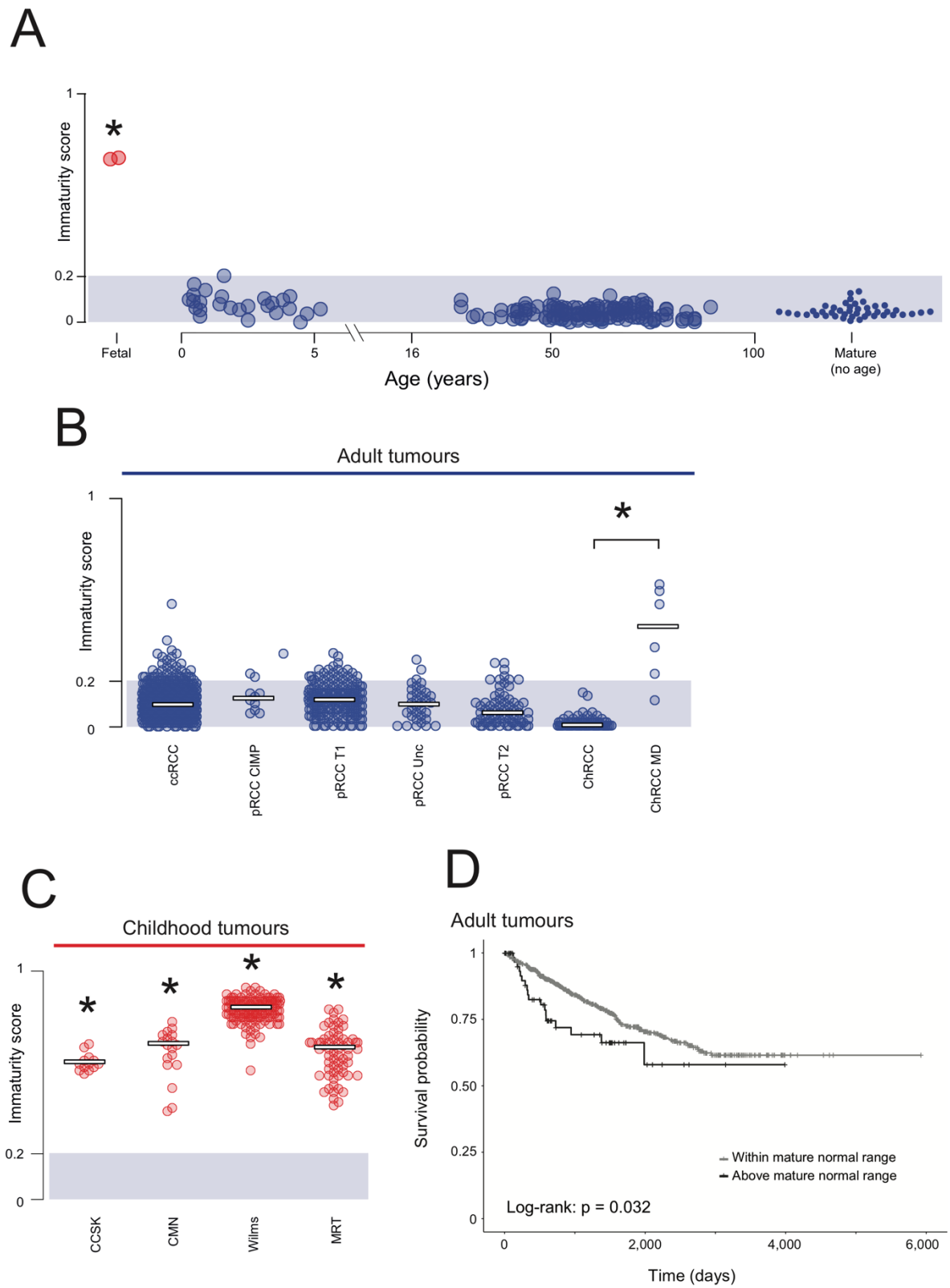

A
Immaturity score
1
0.2
0
*
Fetal
0
5
16
50
100
Mature
(no age)
Age (years)
B
Adult tumours
Immaturity score
1
0.2
0
*
ccRCC
pRCC CIMP
pRCC T1
pRCC Unc
pRCC T2
ChRCC
ChRCC MD
C
Childhood tumours
Immaturity score
1
0.2
0
*
*
*
*
CCSK
CMN
Wilms
MRT
D
Adult tumours
Survival probability
1
0.75
0.50
0.25
0
Within mature normal range
Above mature normal range
Log-rank: p = 0.032
0
2,000
4,000
6,000
Time (days)

Figure 2 – Immaturity score

A. **Immaturity score for normal renal bulk transcriptomes:** For 201 bulk transcriptomes derived from normal kidney biopsies across the human life span, an immaturity score was calculated. This score was calculated by fitting each bulk transcriptome using a combined reference set consisting of signals from all cells in the mature and fetal kidney. The immaturity score is then the total normalized signal contribution from fetal kidney in each bulk transcriptome (shown on the y-axis). Samples are shown by age when known (x-axis), with samples for which no age information was available shown on the right. The shaded blue area indicates the range of maturity scores across all normal tissue post-natal transcriptomes. Bulk fetal transcriptomes are shown on the left in red, with a star indicating that their maturity score is significantly higher than that of the normal samples ($p=0.015$, Wilcoxon rank sum test). Sample specific immaturity scores can be found in **Supplementary Table 6**.
B. **Immaturity score for adult renal tumor transcriptomes:** The immaturity score is calculated as in **A.** for 853 adult renal tumors. The normal immaturity score range is shown by the blue shaded region. The metabolically divergent subtype of Chromophobe renal cell carcinomas are shown separately from other chromophobe RCCs. These metabolically divergent tumors have a significantly different maturity score ($p<10^{-4}$, Wilcoxon rank sum test). Sample specific immaturity scores can be found in **Supplementary Table 6**.
C. **Immaturity score for childhood renal tumor transcriptomes**: The immaturity score is calculated as in **A.** for 287 childhood renal tumors. The normal immaturity score range is shown by the blue shaded region. Each type of childhood tumor had a significantly different maturity score than post-natal normal tissue kidneys ($p<10^{-4}$, Wilcoxon rank sum test). Sample specific immaturity scores can be found in **Supplementary Table 6**.
D. **Survival difference for immature adult renal tumors:** A Kaplan-Meier curve showing the difference in survival probability for adult renal tumors with immaturity scores in the post-natal normal tissue range (grey lines/points) and those above (black lines/points). Note that metabolically divergent chromophobe tumors were excluded from this analysis. The survival probability is shown on the y-axis and the x-

axis shows the survival time from initial diagnosis in days. The difference between the two groups was found to be significant (p=0.0015, Log-rank test).

Figure 3

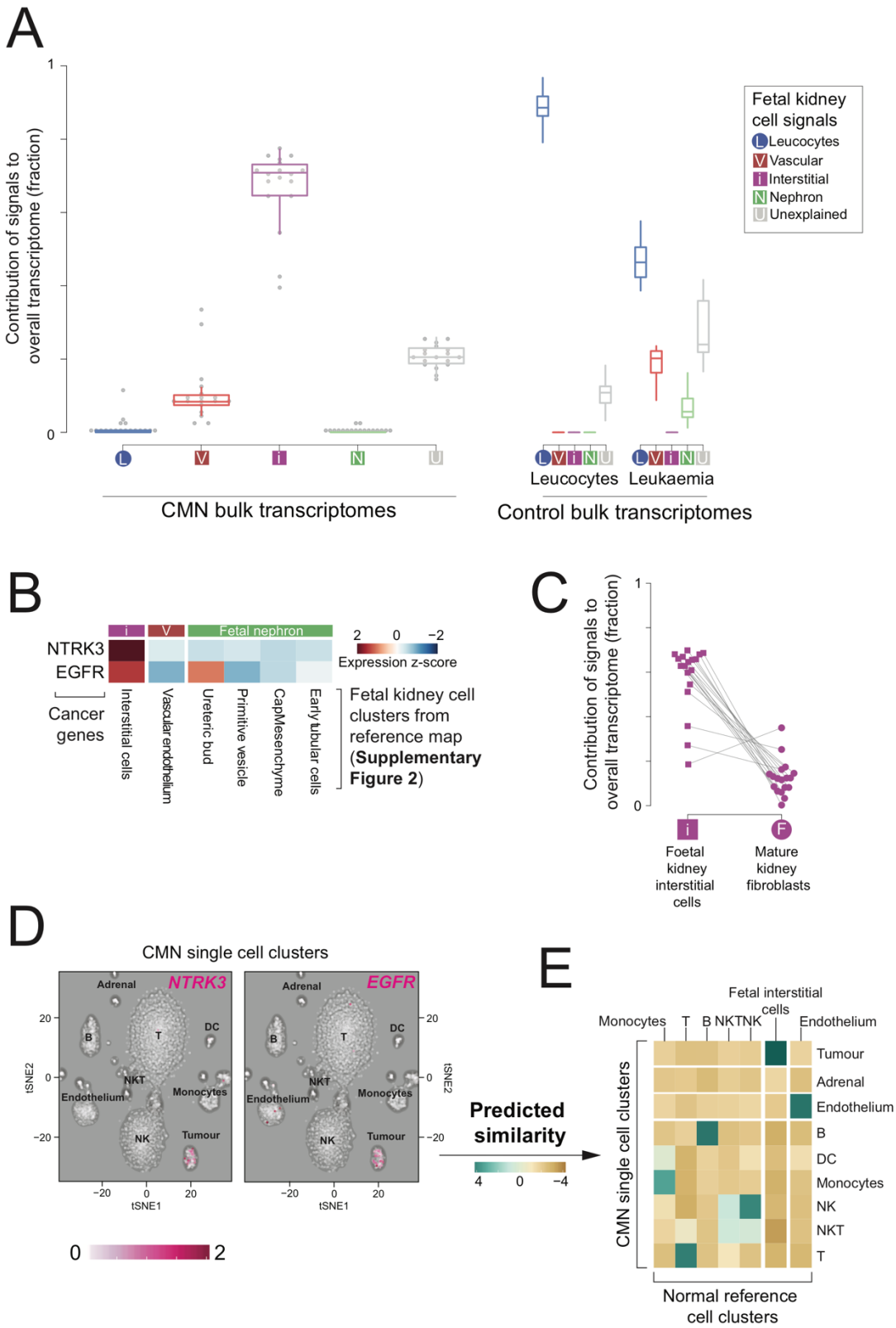

A
Contribution of signals to overall transcriptome (fraction)
1
0
Fetal kidney cell signals
Leucocytes
Vascular
Interstitial
Nephron
Unexplained
Leucocytes
Leukaemia
CMN bulk transcriptomes
Control bulk transcriptomes
B
Fetal nephron
NTRK3
EGFR
Expression z-score
Cancer genes
Interstitial cells
Vascular endothelium
Ureteric bud
Primitive vesicle
CapMesenchyme
Early tubular cells
Fetal kidney cell clusters from reference map (Supplementary Figure 2)
C
Contribution of signals to overall transcriptome (fraction)
Foetal kidney interstitial cells
Mature kidney fibroblasts
D
CMN single cell clusters
NTRK3
EGFR
Adrenal
B
T
DC
NKT
Endothelium
Monocytes
NK
Tumour
tSNE1
tSNE2
Predicted similarity
E
Fetal interstitial cells
Monocytes
T
B
NK
T
NK
Endothelium
Tumour
Adrenal
Endothelium
B
DC
Monocytes
NK
NKT
T
CMN single cell clusters
Normal reference cell clusters

Figure 3 – Congenital mesoblastic nephroma

A. **Composition of bulk CMNs:** The relative contribution of single cell derived signals from fetal kidney in explaining the bulk transcriptomes of 18 congenital mesoblastic nephromas (CMNs) along with control leucocyte and ALL populations. The relative contribution of each signal to each bulk RNA-seq sample is shown by the y-axis. Each signal/sample combination is represented by a single point and the distribution of relative signal contributions to the bulk transcriptomes are summarized with boxplots. Each signal type is labelled with an abbreviation and colored as shown by the legend on the right. Signals are marked with a square for fetal kidney and circle for mature. CMN samples are shown in the block on the left, while the two groups of control samples are shown on the right. The samples marked "Leukocytes" are bulk transcriptomes from flow sorted leukocytes, while the samples marked "Leukemia" represent B-precursor acute lymphoblastic leukemia. Raw signal contributions are listed in **Supplementary Table 7**.
B. **Expression of CMN cancer genes in fetal kidney:** Expression of CMN driver genes (rows) in reference fetal kidney single cell RNA-seq populations (columns). The data has been scaled to have mean 0 and a standard deviation of 1 in each row (i.e., z-transformed).
C. **Comparing fetal versus mature fibroblast signals in CMN bulk transcriptomes:** All 18 CMN bulk transcriptomes were analyzed using a reference signal set including both fetal kidney interstitial cells and the equivalent population in mature kidney. This figure shows the comparison of their inferred contribution to each transcriptome for each sample (y-axis), with lines joining points representing the same sample.
D. **Expression of CMN marker genes:** tSNE map of single cell transcriptomes of 4,416 cells derived from a CMN biopsy. Cells belonging to the same cluster are indicated by contours and are labelled by the cell type they represent (see **Methods**, **Supplementary Table 8**). The normalized expression of *NTRK3* (left) and *EGFR* (right), genes specific to CMN tumor cells, is shown by the color scheme for each point. Abbreviations: B = B cell; T = T cell; DC = dendritic cell; NK = NK cell; NKT = NKT cell. Justification for this annotation are given in **Supplementary Table 9**.

E. **Comparison of single cell CMN to fetal kidney:** Comparison of clusters of cells for which single cell transcriptomes were obtained from the CMN biopsy (rows) with fetal kidney and leucocyte reference populations (columns). For each CMN cluster/reference population pair a log-similarity score was calculated using logistic regression (see Methods). Positive log-similarity scores represent a high probability of similarity between the reference and test cluster. Abbreviations, as per 3D.

Figure 4

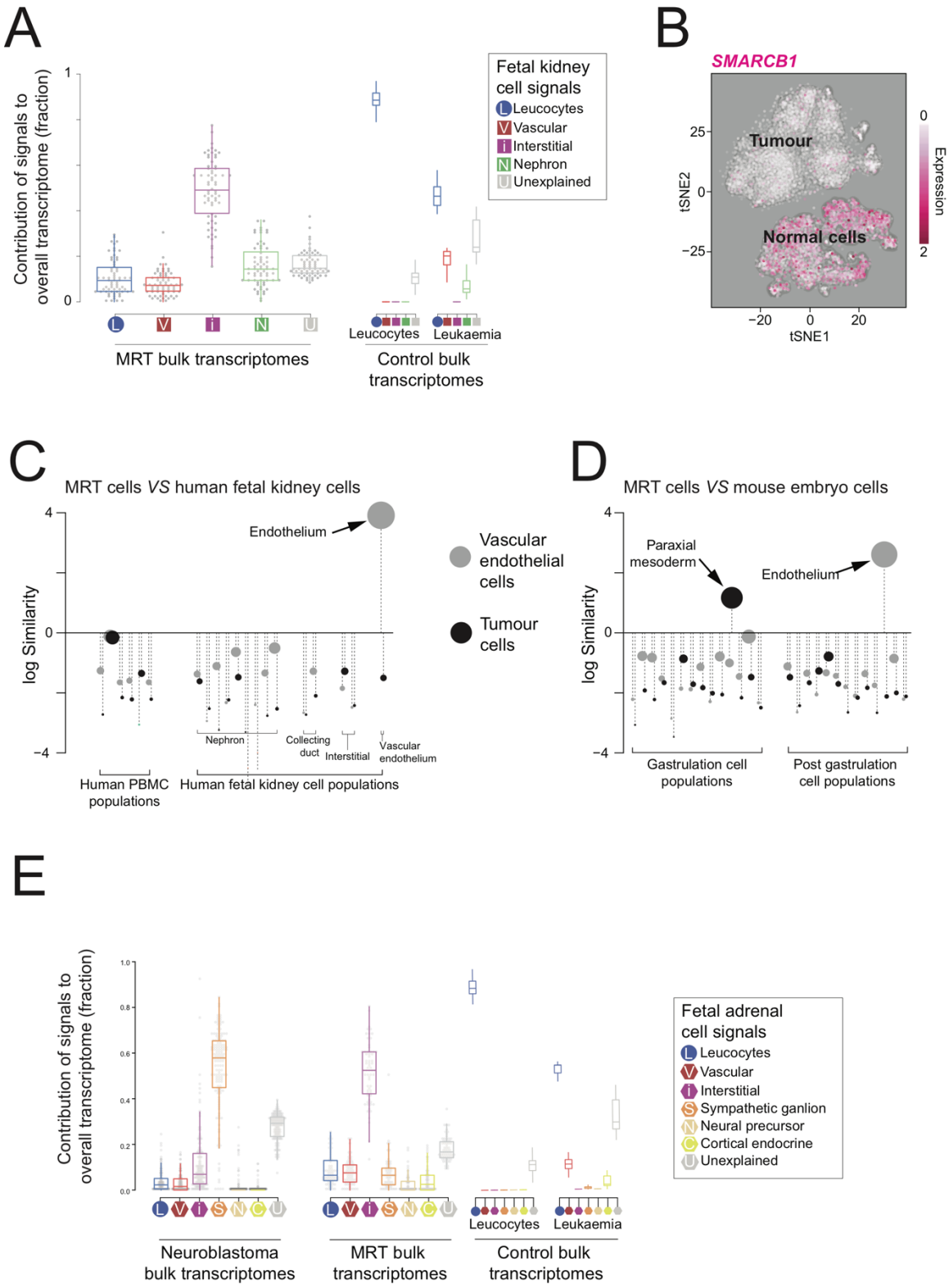

A
Contribution of signals to overall transcriptome (fraction)
1
0
Fetal kidney cell signals
Leucocytes
Vascular
Interstitial
Nephron
Unexplained
Leucocytes
Leukaemia
MRT bulk transcriptomes
Control bulk transcriptomes
B
SMARCB1
Tumour
Normal cells
tSNE2
tSNE1
25
0
−25
−20
0
20
Expression
0
2
C
MRT cells VS human fetal kidney cells
log Similarity
4
0
−4
Endothelium
Vascular endothelial cells
Tumour cells
Nephron
Collecting duct
Interstitial
Vascular endothelium
Human PBMC populations
Human fetal kidney cell populations
D
MRT cells VS mouse embryo cells
log Similarity
4
0
−4
Paraxial mesoderm
Endothelium
Gastrulation cell populations
Post gastrulation cell populations
E
Contribution of signals to overall transcriptome (fraction)
1.0
0.8
0.6
0.4
0.2
0.0
Fetal adrenal cell signals
Leucocytes
Vascular
Interstitial
Sympathetic ganlion
Neural precursor
Cortical endocrine
Unexplained
Leucocytes
Leukaemia
Neuroblastoma bulk transcriptomes
MRT bulk transcriptomes
Control bulk transcriptomes

Figure 4 – Malignant rhabdoid tumors

A. **Bulk MRTs compared to fetal kidney:** The relative contribution of single cell derived signals from fetal kidney in explaining the bulk transcriptomes of 65 malignant rhabdoid tumors (MRTs) along with control leukocyte and ALL populations. The relative contribution of each signal to each bulk RNA-seq sample is shown by the y-axis. Each signal/sample combination is represented by a single point and the distribution of relative signal contributions to the bulk transcriptomes are summarized with boxplots. Each signal type is labelled with an abbreviation and colored as shown by the legend on the right. Signals are marked with a square for fetal kidney and circle for mature. MRT samples are shown in the block on the left, while the two groups of control samples are shown on the right. The samples marked "Leucocytes" are bulk transcriptomes from flow sorted leucocytes, while the samples marked "Leukemia" represent B-precursor acute lymphoblastic leukemia. Raw signal contributions are listed in **Supplementary Table 10**.
B. **Expression of MRT driver gene in MRT organoids:** Each dot represents a single cell transcriptome from organoids derived from MRT tumor cells, or normal distal kidney tubules. Cells have been placed in 2-dimensions using the tSNE dimension reduction technique and points colored by the gene *SMARCB1* (INI1), a disease defining alteration of MRT. Cell specific information is given in **Supplementary Table 11**.
C. **Log similarity of single cell MRT cells to fetal kidney:** Comparison of MRT tumor organoid cells (black dots), and mature vascular cells from adult kidney (grey dots) to PBMC, and fetal kidney single cell defined references. Similarity scores were calculated using logistic regression trained on fetal kidney plus PBMC single cell transcriptomes. The predicted similarity of the tumor single cells for each of the reference populations shown on the x-axis is then calculated using this trained logistic regression model. Vascular endothelial cells are included as a control. The y-axis shows this similarity score on a log scale, with high values representing a strong similarity and negative values strong dis-similarity. Points are scaled in proportion to linear similarities. Each dot represents an average of the log-similarity across all cells of the type indicated by the color scheme (i.e., tumor or vascular). Those reference

populations for which either the vascular or tumor cells have a high similarity are labelled with arrows.

D. **Log similarity of single cell MRT cells to developing mouse:** The same as in **C** except the comparison has been made to a reference derived from single cell mouse data, covering the period of development immediately following gastrulation. Only mouse/human orthologs genes were considered in the comparison. Those reference populations for which either the vascular or tumor cells have a high similarity are labelled with arrows.

E. **Bulk MRTs compared to fetal adrenal:** The relative contribution of single cell derived signals from fetal adrenal glands in explaining the bulk transcriptomes of 65 malignant rhabdoid tumors (MRTs), and 142 neuroblastomas along with control leukocyte and ALL populations. The relative contribution of each signal to each bulk RNA-seq sample is shown by the y-axis. Each signal/sample combination is represented by a single point and the distribution of relative signal contributions to the bulk transcriptomes are summarized with boxplots. Each signal type is labelled with an abbreviation and colored as shown by the legend on the right. Signals are marked with a square for fetal kidney, circle for mature, and hexagon for fetal adrenal. MRT, and neuroblastoma samples are shown in the blocks on the left, while the two groups of control samples are shown on the right. The samples marked "Leukocyte" are bulk transcriptomes from flow sorted leukocytes, while the samples marked "ALL" represent B-precursor acute lymphoblastic leukemia. Raw signal contributions are listed in **Supplementary Table 12**.

Figure 5

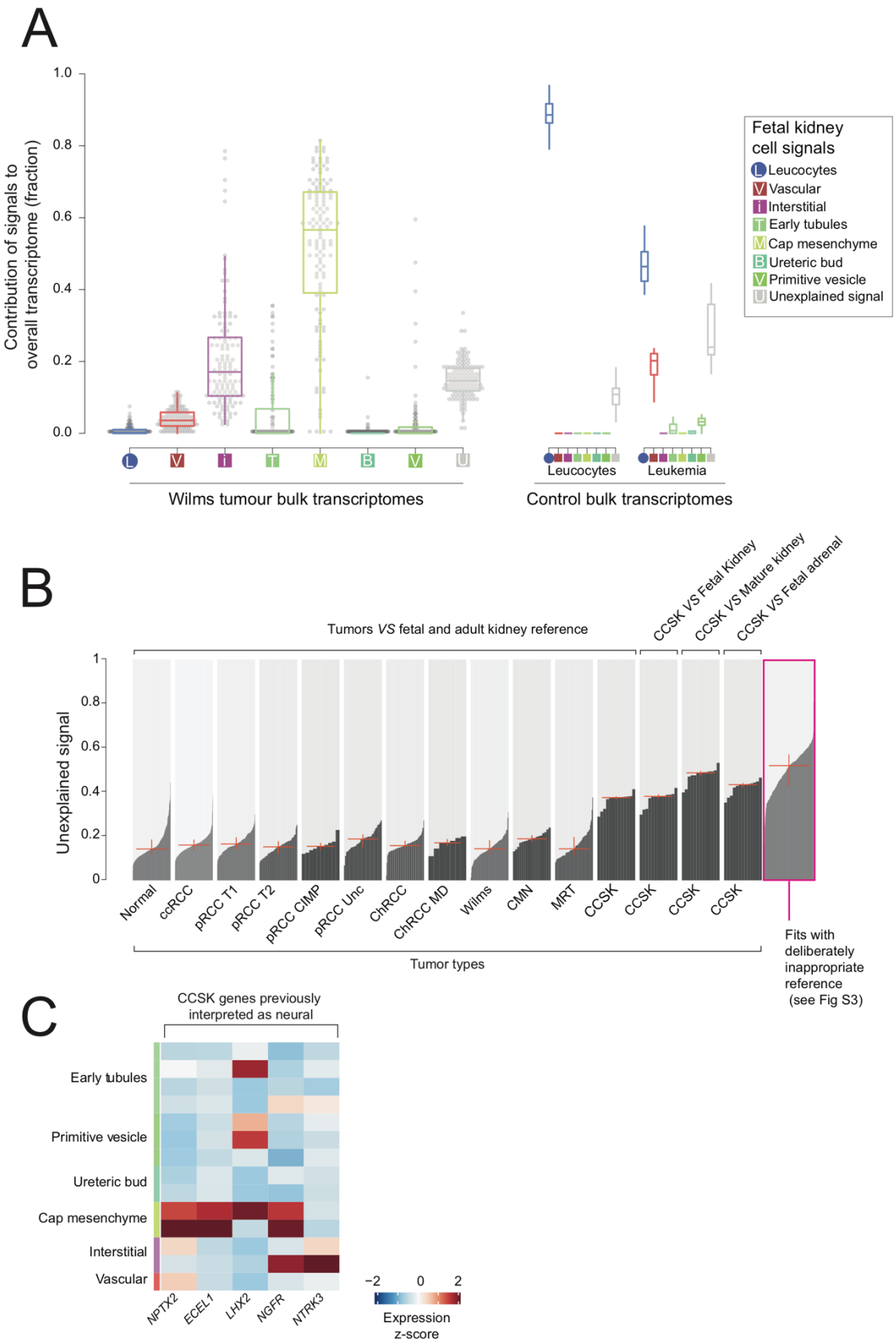

A
Contribution of signals to overall transcriptome (fraction)
1.0
0.8
0.6
0.4
0.2
0.0
Fetal kidney cell signals
Leucocytes
Vascular
Interstitial
Early tubules
Cap mesenchyme
Ureteric bud
Primitive vesicle
Unexplained signal
Leucocytes
Leukemia
Wilms tumour bulk transcriptomes
Control bulk transcriptomes
B
Tumors VS fetal and adult kidney reference
CCSK VS Fetal Kidney
CCSK VS Mature kidney
CCSK VS Fetal adrenal
Unexplained signal
Normal
ccRCC
pRCC T1
pRCC T2
pRCC CIMP
pRCC Unc
ChRCC
ChRCC MD
Wilms
CMN
MRT
CCSK
CCSK
CCSK
CCSK
Tumor types
Fits with deliberately inappropriate reference (see Fig S3)
C
CCSK genes previously interpreted as neural
Early tubules
Primitive vesicle
Ureteric bud
Cap mesenchyme
Interstitial
Vascular
NPTX2
ECEL1
LHX2
NGFR
NTRK3
−2
0
2
Expression z-score

Figure 5 – Wilms tumor and clear cell sarcoma of the kidney

A. **Bulk Wilms tumor compared to fetal kidney:** The relative contribution of single cell derived signals from the fetal kidney in explaining the bulk transcriptomes of 120 nephroblastomas (Wilms) along with control leucocyte and ALL populations. The signal intensity assigned to each sample is shown by the y-axis, normalized so that the sum of all signal contributions is 1 for each sample. The relative contribution of each signal to each bulk RNA-seq sample is shown by the y-axis. Each signal/sample combination is represented by a single point and the distribution of relative signal contributions to the bulk transcriptomes are summarized with boxplots. Each signal type is labelled with an abbreviation and colored as shown by the legend on the right. Signals are marked with a square for fetal kidney and circle for mature. Wilms tumor samples are shown in the block on the left, while the two groups of control samples are shown on the right. The samples marked "Leucocytes" are bulk transcriptomes from flow sorted leukocytes, while the samples marked "Leukemia" represent B-precursor acute lymphoblastic leukemia. Raw signal contributions are listed in **Supplementary Table 13**.
B. **Comparison of unexplained signal contribution to CCSKs and other tumor types:** Comparison of the fraction of the bulk transcriptomes attributed to the "unexplained signal" in clear cell sarcoma of the kidney samples (CCSK) and other groups of samples. For each group of samples, the unexplained signal is calculated for each sample individually, using the reference set of signals given at the top of the plot (e.g., fetal kidney). The unexplained signal fractions are then shown by black bars, sorted in increasing order, with the red horizontal line showing the median value and the vertical line the range between the 25th and 75th percentiles. CCSK samples were fitted using 4 different reference sets (fetal and mature kidney, fetal kidney only, mature kidney only, and fetal adrenal). The final group on the right, represents samples fitted using inappropriate references (**Supplementary Figure 3**). This population serves as a calibration of the expected level of unexplained signal when the bulk transcriptome is not explained by any of the provided reference signals. The full CCSK fit is shown in **Supplementary Figure 9** and **Supplementary Table 14**.

C. **Expression of neural genes in fetal kidney:** Expression of genes identified as "neural" and upregulated in CCSK compared to normal kidney in fetal kidney populations. Normalized expression values are averaged within each population of cells for each gene and these values are then transformed to have a mean of 0 and a standard deviation of 1 in each column (i.e., they are z-transformed). Rows represent different populations of cell in the reference fetal kidney, columns represent genes of interest.

Figure 6

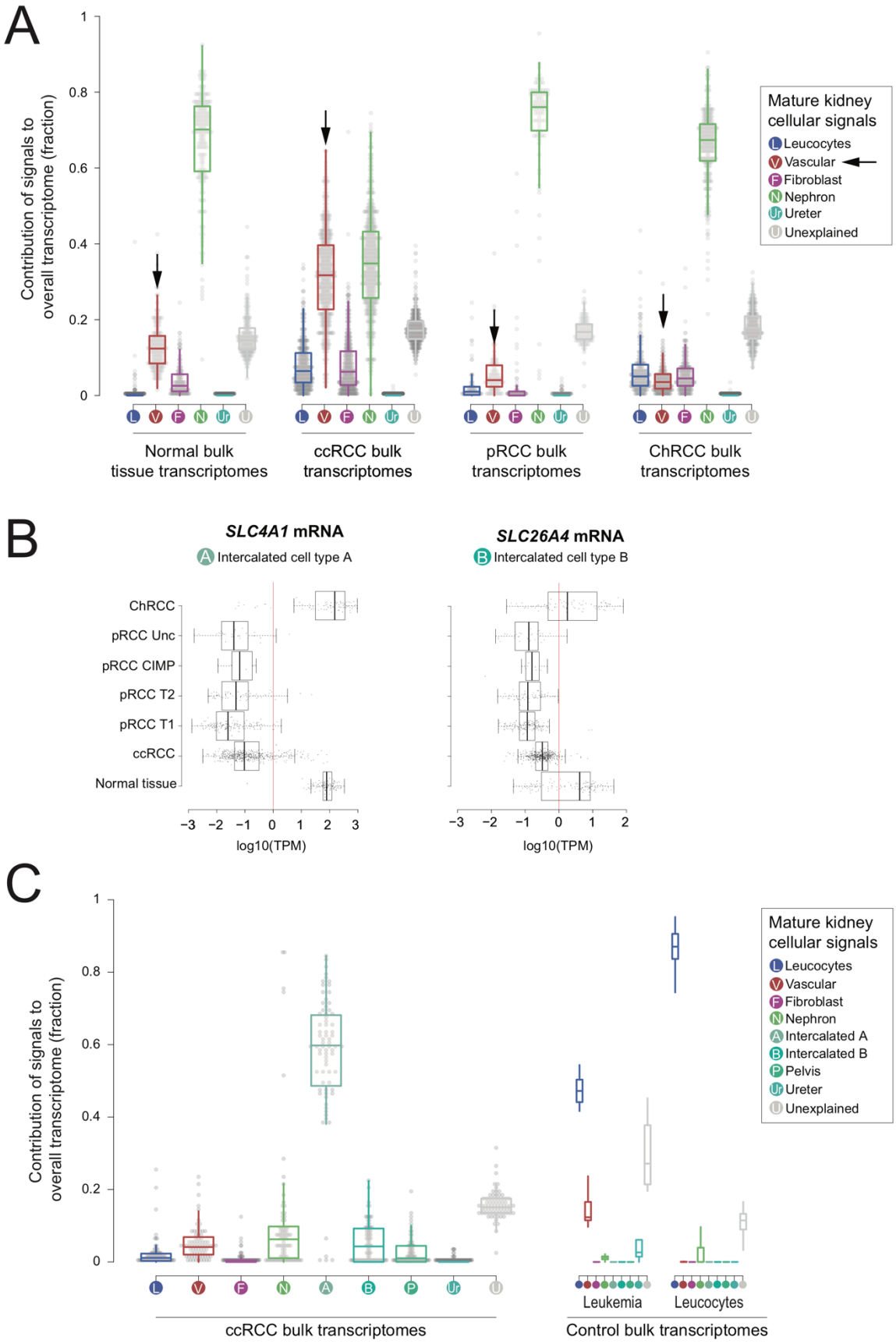

A
Contribution of signals to overall transcriptome (fraction)
Mature kidney cellular signals
Leucocytes
Vascular
Fibroblast
Nephron
Ureter
Unexplained
Normal bulk tissue transcriptomes
ccRCC bulk transcriptomes
pRCC bulk transcriptomes
ChRCC bulk transcriptomes
B
SLC4A1 mRNA
Intercalated cell type A
SLC26A4 mRNA
Intercalated cell type B
ChRCC
pRCC Unc
pRCC CIMP
pRCC T2
pRCC T1
ccRCC
Normal tissue
log10(TPM)
C
Contribution of signals to overall transcriptome (fraction)
Mature kidney cellular signals
Leucocytes
Vascular
Fibroblast
Nephron
Intercalated A
Intercalated B
Pelvis
Ureter
Unexplained
Leukemia
Leucocytes
ccRCC bulk transcriptomes
Control bulk transcriptomes

Figure 6 – Adult kidney tumors

A. **Bulk renal cell carcinomas compared to mature kidney:** The relative contribution of single cell derived signals from mature kidneys in explaining the bulk transcriptomes of 171 normal kidney biopsies, 500 clear cell renal cell carcinomas (ccRCC), 274 papillary renal cell carcinomas (pRCC), and 81 chromophobe renal cell carcinomas (ChRCC). The relative contribution of each signal to each bulk RNA-seq sample is shown by the y-axis. Each signal/sample combination is represented by a single point and the distribution of relative signal contributions to the bulk transcriptomes are summarized with boxplots. Each signal type is labelled with an abbreviation and colored as shown by the legend on the right. Signals are marked with a circle for mature kidneys. The contribution to the bulk transcriptomes for all leucocyte, vascular, and nephron signals are aggregated together. Raw signal contributions are listed in **Supplementary Table 15**.
B. **Expression of markers of Type A and B intercalated cells in bulk transcriptomes:** Boxplots showing expression of canonical markers of Type A (left) and B (right) intercalated cells in different populations of bulk kidney tumor and normal transcriptomes. Expression values are normalized to transcripts per million (TPM) and log10 transformed and shown on the x-axis, with each point representing a sample and a boxplot showing the distribution of expression values for that sample type. A red line marks 0 on the log transformed expression scale.
C. **Bulk chromophobe renal cell carcinomas compared to mature kidney:** The relative contribution of single cell derived signals from the mature kidney in explaining the bulk transcriptomes of 81 chromophobe renal cell carcinomas (ChRCC), along with control leucocyte and ALL populations. The relative contribution of each signal to each bulk RNA-seq sample is shown by the y-axis. Each signal/sample combination is represented by a single point and the distribution of relative signal contributions to the bulk transcriptomes are summarized with boxplots. Each signal type is labelled with an abbreviation and colored as shown by the legend on the right. Signals are marked with a circle for mature kidneys. ChRCC samples are shown in the blocks on the left, while the two groups of control samples are shown on the right. The samples marked “Leucocyte” are bulk transcriptomes from flow sorted leucocytes,

while the samples marked “Leukemia” represent B-precursor acute lymphoblastic leukemia. Raw signal contributions are listed in **Supplementary Table 16**.

Figure 7

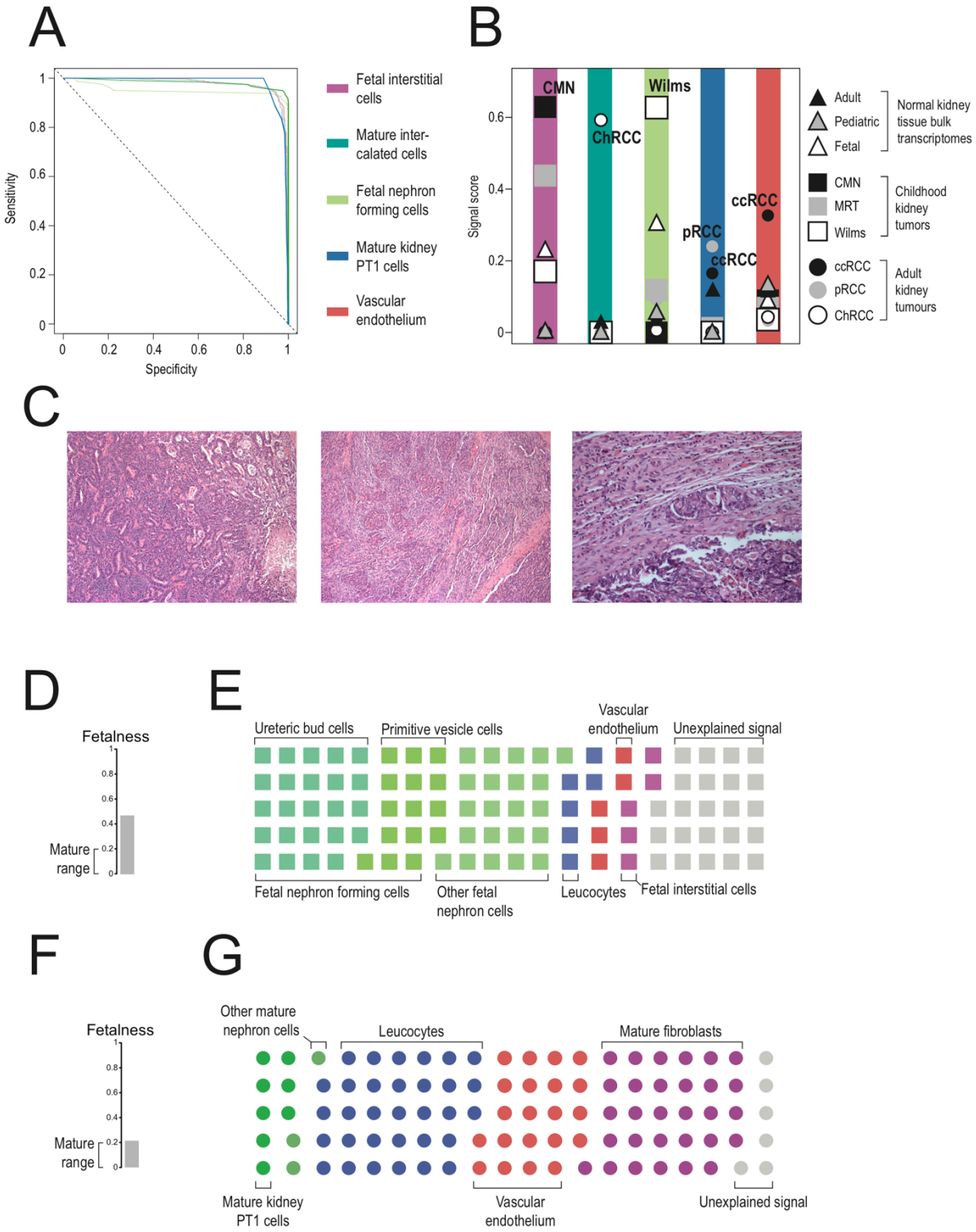

A
Sensitivity
Specificity
Fetal interstitial cells
Mature inter-calated cells
Fetal nephron forming cells
Mature kidney PT1 cells
Vascular endothelium
B
Signal score
CMN
ChRCC
Wilms
pRCC
ccRCC
ccRCC
Adult
Pediatric
Fetal
Normal kidney tissue bulk transcriptomes
CMN
MRT
Wilms
Childhood kidney tumors
ccRCC
pRCC
ChRCC
Adult kidney tumours
C
D
Fetalness
Mature range
E
Ureteric bud cells
Primitive vesicle cells
Vascular endothelium
Unexplained signal
Fetal nephron forming cells
Other fetal nephron cells
Leucocytes
Fetal interstitial cells
F
Fetalness
Mature range
G
Other mature nephron cells
Leucocytes
Mature fibroblasts
Mature kidney PT1 cells
Vascular endothelium
Unexplained signal

Figure 7 – Clinical utility of cellular signal analysis

A. **Sensitivity/Specificity of signals in classifying tumor types:** Curves showing the sensitivity and specificity of using the scores defined by the color scheme to classify tumors by type at different cut-offs. The different score and tumor type pairs are: fetal interstitial cells and CMN (light blue), intercalated cells and ChRCCs (dark blue), the developing nephron and Nephroblastoma (light green), PT1 and ccRCCs/pRCCs (dark green), and mature vascular and ccRCCs (red).
B. **Median reference contribution by tumor type:** Each point represents the median score for the group of samples indicated by the combination of shape and shading, as explained by the legend on the right, for the score type specified by the background shading. Score types are the same as in **A**. Sample level data relating to this Figure is shown in Supplementary Figure 12.
C. **Histology image of unclassified childhood renal tumor:** The tumor mostly compromised pleomorphic epithelioid cells that formed tubules, papillae, glands and nests, as well as more solid areas with spindled cells and clefting similar to that of synovial sarcoma. Patchy tumor necrosis was apparent. Some areas showed smaller, more uniform cells lining narrow tubular structures, resembling adenomatous perilobar nephrogenic rests. Overall, the morphology and ancillary tests were inconclusive.
D. **Immaturity score for unclassified childhood renal tumor:** The immaturity score for the unclassified childhood renal tumor, calculated as in **Figure 2**. The range of immaturity scores found in the normal mature kidney are indicated with the label on the left.
E. **Summary of signal contribution from fetal and mature kidney to unclassified childhood renal tumor:** The relative contribution of single cell derived signals from the fetal kidney in explaining the bulk transcriptomes of the unclassified tumor from **C**. The 100 squares are colored so the fraction of squares of each color matches the fractional contribution of each fetal kidney signal to explaining the bulk transcriptome of this sample. The labels above and below indicate what each colour represents. Raw signal contributions for this sample are listed in **Supplementary Table 17**.

F. **Immaturity score for childhood renal cell carcinoma:** The immaturity score for a childhood renal cell carcinoma, calculated as in **Figure 2**. The range of immaturity scores found in normal mature kidney are indicated with the label on the left.

G. **Summary of signal contribution from fetal and mature kidney to childhood renal cell carcinoma:** As in **D/E** but for a transcriptome derived from renal cell carcinoma fit using a mature kidney signal set. Raw signal contributions for this sample are listed in **Supplementary Table 17**.

References

1 Blum, A., Wang, P. & Zenklusen, J. C. SnapShot: TCGA-Analyzed Tumors. *Cell* **173**, 530, doi:10.1016/j.cell.2018.03.059 (2018).

2 International Cancer Genome, C. *et al.* International network of cancer genome projects. *Nature* **464**, 993-998, doi:10.1038/nature08987 (2010).

3 Lindgren, D. *et al.* Cell-Type-Specific Gene Programs of the Normal Human Nephron Define Kidney Cancer Subtypes. *Cell Rep* **20**, 1476-1489, doi:10.1016/j.celrep.2017.07.043 (2017).

4 Chen, F. *et al.* Multilevel Genomics-Based Taxonomy of Renal Cell Carcinoma. *Cell Rep* **14**, 2476-2489, doi:10.1016/j.celrep.2016.02.024 (2016).

5 Young, M. D. *et al.* Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594-599, doi:10.1126/science.aat1699 (2018).

6 Rini, B. I. & Small, E. J. Biology and clinical development of vascular endothelial growth factor-targeted therapy in renal cell carcinoma. *J Clin Oncol* **23**, 1028-1043, doi:10.1200/JCO.2005.01.186 (2005).

7 Behjati, S., Lindsay, S., Teichmann, S. A. & Haniffa, M. Mapping human development at single-cell resolution. *Development* **145**, doi:10.1242/dev.152561 (2018).

8 Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, doi:10.7554/eLife.27041 (2017).

9 Lake, B. B. *et al.* A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. *Nat Commun* **10**, 2832, doi:10.1038/s41467-019-10861-2 (2019).

10 Arazi, A. *et al.* The immune cell landscape in kidneys of patients with lupus nephritis. *Nat Immunol* **20**, 902-914, doi:10.1038/s41590-019-0398-x (2019).

11 Hochane, M. *et al.* Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development. *PLoS Biol* **17**, e3000152, doi:10.1371/journal.pbio.3000152 (2019).

12 Wang, P. *et al.* Dissecting the Global Dynamic Molecular Profiles of Human Fetal Kidney Development by Single-Cell RNA Sequencing. *Cell Rep* **24**, 3554-3567 e3553, doi:10.1016/j.celrep.2018.08.056 (2018).

13 Menon, R. *et al.* Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney. *Development* **145**, doi:10.1242/dev.164038 (2018).

14 Kang, K. *et al.* A novel computational complete deconvolution method using RNA-seq data. *bioRxiv*, 496596, doi:10.1101/496596 (2018).

15 Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* **10**, 380, doi:10.1038/s41467-018-08023-x (2019).

16 Ricketts, C. J. *et al.* The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep* **23**, 313-326 e315, doi:10.1016/j.celrep.2018.03.075 (2018).

17 Vujanic, G. M. & Charles, A. K. Renal tumours of childhood: an update. *Pathology* **40**, 217-227, doi:10.1080/00313020701813776 (2008).

18 Wegert, J. *et al.* Recurrent intragenic rearrangements of EGFR and BRAF in soft tissue tumors of infants. *Nat Commun* **9**, 2378, doi:10.1038/s41467-018-04650-6 (2018).

19 Chun, H. E. *et al.* Genome-Wide Profiles of Extra-cranial Malignant Rhabdoid Tumors Reveal Heterogeneity and Dysregulated Developmental Pathways. *Cancer Cell* **29**, 394-406, doi:10.1016/j.ccell.2016.02.009 (2016).
20 Pijuan-Sala, B. *et al.* A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490-495, doi:10.1038/s41586-019-0933-9 (2019).
21 Dome, J. S. *et al.* Advances in Wilms Tumor Treatment and Biology: Progress Through International Collaboration. *J Clin Oncol* **33**, 2999-3007, doi:10.1200/JCO.2015.62.1888 (2015).
22 Iliopoulos, O., Levy, A. P., Jiang, C., Kaelin, W. G., Jr. & Goldberg, M. A. Negative regulation of hypoxia-inducible genes by the von Hippel-Lindau protein. *Proc Natl Acad Sci U S A* **93**, 10595-10599, doi:10.1073/pnas.93.20.10595 (1996).
23 Tannir, N. M. *et al.* A phase 2 trial of sunitinib in patients with advanced non-clear cell renal cell carcinoma. *Eur Urol* **62**, 1013-1019, doi:10.1016/j.eururo.2012.06.043 (2012).
24 Davis, C. F. *et al.* The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* **26**, 319-330, doi:10.1016/j.ccr.2014.07.014 (2014).
25 Welsh-Bacic, D., Nowik, M., Kaissling, B. & Wagner, C. A. Proliferation of acid-secretory cells in the kidney during adaptive remodelling of the collecting duct. *PLoS One* **6**, e25240, doi:10.1371/journal.pone.0025240 (2011).