HHS Public Access

Author manuscript

Kidney Int. Author manuscript; available in PMC 2015 May 01.

Published in final edited form as:

Kidney Int. 2014 November ; 86(5): 1049–1055. doi:10.1038/ki.2014.104.

Validating single cell genomics for the study of renal development

Sanjay Jain, MD, PhD[1,2,#,*], **Michiel J. Noordam, PhD**[3,#], **Masato Hoshi, MD, PhD**[1,#], **Francesco L. Vallania, PhD**[3,4], and **Donald F. Conrad, PhD**[3,4,#,*]

[1]Department of Internal Medicine (Renal Division), Washington University School of Medicine

[2]Department of Pathology & Immunology, Washington University School of Medicine

[3]Department of Genetics, Washington University School of Medicine

[4]Center for Genome Sciences and Systems Biology, Washington University School of Medicine

Abstract

Single-cell genomics will enable studies of the earliest events in kidney development, although it is unclear if existing technologies are mature enough to generate accurate and reproducible data on kidney progenitors. Here we designed a pilot study to validate a high-throughput assay to measure the expression levels of key regulators of kidney development in single cells isolated from embryonic mice. Our experiment produced 4,608 expression measurements of 22 genes, made in small cell pools and 28 single cells purified from the RET-positive ureteric bud. There were remarkable levels of concordance with expression data generated by traditional microarray analysis on bulk ureteric bud tissue with the correlation between our average single cell measurements and GUDMAP measurements for each gene of 0.82-0.85. Nonetheless, a major motivation for single cell technology is to uncover dynamic biology hidden in population means. There was evidence for extensive and surprising variation in expression of Wnt11 and Etv5, both downstream targets of activated RET. The variation for all genes in the study was strongly consistent with burst-like promoter kinetics. Thus, our results can inform the design of future single-cell experiments, which are poised to provide important insights into kidney development and disease.

Keywords

Single cell; ureteric bud; kidney; Ret

[*]Corresponding Authors: Donald F. Conrad, Department of Genetics, Washington University School of Medicine, Campus Box 8232, St. Louis MO, 63110, USA, dconrad@genetics.wustl.edu, Phone: +1 314-362-4379, FAX: +1 314-362-7855, Sanjay Jain, Department of Internal Medicine, Washington University School of Medicine, Campus Box 8126, St. Louis MO, 63110, USA, jains@dom.wustl.edu, Phone: +1 314-454-8728, FAX: +1 314-454-7735.

[#]contributed equally

Disclosure: We disclose that there are no conflicts of financial interest.

The development of each kidney depends on precise spatiotemporal interactions between the ureteric bud (UB) and the metanephric mesenchyme beginning at about five weeks of gestation in humans and 10.5 days post-coitum (e10.5) in mice[1]. These interactions trigger a highly coordinated developmental program between the derivatives of epithelial, mesenchymal and endothelial progenitors that ultimately give rise to the mature kidney. In order to create or sustain a native or synthetic kidney it is therefore essential to create a molecular narrative of kidney development that should include the timing and location of changes in gene expression, chromatin and protein activity and cellular metabolism. The field of genomics is currently exploding with ingenious new techniques to measure these cellular processes, mostly empowered by the availability of low-cost DNA synthesis and high throughput sequencing[2, 3]. Furthermore, a revolution in microfluidics now allows us to apply these genomic techniques to nanoliter reaction volumes, enabling measurements to be made at the level of single cells[4]. This convergence has made it feasible, in principle, to describe the genomic program of kidney development at a single cell level, so that molecular events during early kidney development can be deciphered.

In order to implement and validate single cell genomic technology for the study of individual cells of different lineages during kidney development, we selected the branching ureteric bud cells as our model system. These cells have high expression of the Receptor tyrosine kinase (Ret), a gene that is critical for kidney formation and in specification of the collecting system lineage[5]. Ret is highly expressed in the ureteric epithelium in the UB tip during branching morphogenesis and abnormal Ret expression or activity results in a diverse spectrum of renal malformations [5-8]. We designed a pilot experiment to analyze the expression profiles of 24 kidney development genes from Retlabeled single cells during branching morphogenesis using RET-EGFP reporter mice[9]. We addressed the following questions: (a) are single cell gene expression measurements reproducible for this biological system model, (b) do (average) single UB cell gene expression measurements reproduce match published expression levels measured from pooled UB cells, (c) what is the variability among genes and cells in single UB cell expression levels?

We used FACS to populate 48 wells of a microtiter plate with EGFP-positive cells originating from ten E13.5 metanephroi (**Methods**, Figure 1). In total, 32,020 cells were EGFP-positive, an average of about 3,000 cells per metanephros, comprising 7.9% of the total cell population of each metanephros. Cells were distributed across the 48 wells as follows: 28 single cells, two standard dilution series of (2,4,8,16) cell pools and one standard series of (2,4,8,100) cell pools. Four wells were kept empty as negative controls. cDNA was synthesized in each well and loaded into an integrated microfluidic circuit (IFC). We used qPCR to measure the expression level of 24 genes, selected to represent high (15) or low/absent in the UB (9) based on their known expression in these regions[10] (Supplementary Table 1). Additional rationale for these genes include: 1) Their important roles in kidney development, 2) Several of these interact with RET signaling (GDNF, GFRα1, WNT11, ETV4, ETV5, SPRY1, BMP7, PAX2), 3) They include genes that are known to be expressed exclusively in the UB (example, RET, WNT11, WNT9b), or in the cap mesenchyme (example, GDNF, SIX2, SALL1), or in the stroma (FOXD1), or in the renal vesicle (WNT4) or in more than one of these regions (example, GFRα1 in cap mesenchyme

and UB; BMP7 in UB and renal vesicle; PAX2 in UB and cap mesenchyme). Each primer pair was assayed in duplicate, resulting in 2 × 24 × 48 = 2304 expression measurements per experiment. Finally, the same plate of starting cDNA was assayed in duplicate, by two different scientists, allowing us to quantify the technical measurement error for each experiment. We performed rigorous quality control on each qPCR measurement that subsequently left us with 1562/2304 (68%) measurements that passed in both runs. Two primer sets failed as a whole and all results for these assays were excluded as a result from the QC (**Methods**).

Following within-plate median normalization, we performed analysis of variance (ANOVA) on the Ct values from the single cell measurements and found highly significant variance components associated with well number and primer set, but the variance between plate replicates was not significant (Table 1). The average variance in expression level between cells was over four times greater than within cells (F=8.23 vs F=1.95) with the highest variance, as expected, present among expression levels of different genes (F=156). These results clearly demonstrate that our system allows comparison of gene expression levels between individual cells, and importantly, that there is significant heterogeneity in expression profile among single UB cells.

We next examined the dynamic range and calibration of our qPCR platform. For all 14 genes expressed in UB, we fit linear models relating observed expression (Ct) to cellular copy number using the three dilution series of 2, 4, 8 and 16 cells (Supplementary Figure1). We obtained a mean slope of -1.05, indicating an excellent linear quantitative agreement between cell number and expression level from each gene. Interpolating the single cell measurements with these fitted models produced a mean interpolated cell count of 1.06 across all 14 genes × 28 cells = 392 interpolations. These studies confirmed that our platform is quantitative down to single cell level.

To determine if, at a population level, our single cell measurements agree with pooled cell measurements from the same tissue, we obtained published GUDMAP microarray data for 14 of our genes, generated from expression profiling bulk UB cell pools (**Methods**)[10]. When comparing expression average values of our single cells to GUDMAP data, we observed remarkable concordance: the correlation coefficients between GUDMAP E11.5 UB stalk and our averaged single cell measurements were 0.82 and 0.85, for run 1 and run 2 respectively (Figure 2A). We interpret this finding as clear validation of our single cell system for performing reliable measurements on developing renal cell populations.

One of the primary motivations for performing gene expression analysis at the single cell level is that pooled cell measurements potentially obscure substructure and biology in populations of nominally identical cells, either due to stochastic or programmed variation in expression among cells[11]. Previous single-cell studies of gene expression have suggested that most mammalian promoters display “bursty” kinetics- points of intense transcription separated by long intervals of silence [12, 13]. To determine whether it would also be the case in our population, we examined the cell-to-cell variation in expression levels for each of our genes. We observed a quadratic scaling between mRNA variance and mean, consistent with “burst”-like transcriptional activation (Figure 2B). Next, hierarchical

clustering of the full data demonstrated three groups of genes and two groups of cells (Figure 3). The three gene groups correspond to (a) 8 genes that are robustly expressed across all cells – EGFP, GAPDH, HNF1B, PAX2, GFRA1, GATA3, FGFR2, and WNT9B; (b) 8 genes that are lowly expressed in all cells - FGF9, VEGFR2, SIX2, WNT4, GDNF, SALL1, FOXD1, FOXC2 (non UB genes); (c) 6 genes with coordinated, variable (and nearly bi-modal) expression across cells – RET, WNT11, ETV5, LHX1, BMP7 and SPRY1.

The two cell-type clusters that we observed were totally defined by their expression of genes in group (c). Formal significance testing for the presence of two distinct clades of cells, using multiscale bootstrap resampling on the full set of expression data, did not reject the null hypothesis at a significance of 0.05 (observed $p = 0.26$, Supplementary Figure2). However, support for these 2 distinct cellular groupings increased when clustering only those genes that are expressed in UB (Supplementary Figures3, 4). In 97% of bootstrap samples, all cells with high group (c) expression were more closely clustered to another cell with high group (c) expression than to one of 7 cells with low group (c) expression (cells 1,12,15,16,18,21,24).

We were somewhat surprised to see that one of these two cell populations included apparently EGFP+, RET- cells, as expression of our EGFP reporter is driven by the RET promoter. It is likely that we observe these EGFP+, Ret- cells due to differences in the half-life of EGFP protein, which is about 24 hours, and that of Ret transcripts, which is less than 4 hours [14, 15]. Consistent with the qPCR data, we also observed a small fraction of cells that express EGFP but not Ret (Supplementary Figure 5). Among the 4 other genes in group (c), expression of WNT11 and ETV5 is upregulated by activated RET[6]. However, we observe RET-, ETV5+, WNT11- and RET+, ETV5-, WNT11+ and RET+, ETV5+, WNT11- cell populations that could only be identified through single analysis and missed in pooled analysis. This suggested to us that two cell populations defined by group (c) genes might represent cells in different biological states, perhaps cells that are in transition from an undifferentiated UB tip cell to becoming a stalk cells. The different expression patterns at a single cells level are also supported by expression of Ret and Bmp7 that show high level of expression in most of the cells except a few in which there is low Ret and high Bmp7 (Supplementary Figure3, cells 16 and 24). This is further confirmed by immunohistochemistry where intensity of Ret expression changes from high to low from UB tip to stalk and Bmp7 remains highly expressed (Figure 4A). Immunofluorescence experiments also support that most cells in the UB coexpress high levels of Ret and Wnt11 but there are also rare cells that show high Wnt11 and low Ret (Figure 4B). While the small number of replicates deployed (2 inter-user replicate groups, 29 single cell replicates) limits the assessment of variability of the system our studies provide the necessary validation of single cell analysis on Ret-positive UB cells.

In summary, we have used microfluidic processing for the first time to make accurate measurements of single cell gene expression from Ret-positive UB and validated these by data from standard array-based profiling of bulk cell populations. The data strongly support the conclusion that different combination of genes identify subpools of cells among a large group of cells from the same lineage and the knowledge of these is critical for rationally programming progenitors towards more differentiated cell types. For example, in delineating

the early molecular events that underlie differentiation of the nascent UB tip cell to mature collecting duct cells. Even greater potential for this line of research will be realized with scaling up single cell analysis to profile entire transcriptomes with RNAseq, and to use the resulting datasets to build better computational models for the molecular processes underlying kidney development. Our results provide the necessary proof of concept towards this goal.

Materials and Methods I (for main letter)

Isolation of cells from ureteric bud

All animal studies were conducted in accordance with Institution approved protocols. E13.5 embryonic metanephroi were dissected from five Ret-EGFP embryos (two from each, total 10 metanephroi) and collected in 1.5ml tube. Collected metanephroi were suspended with Trypsin-EDTA solution (Sigma, T3924) with 200μg/ml DNase I (Sigma) and incubated for 10 min at 37 °C. After the incubation the specimens were triturated with a P-200 pipette and rinsed once with DMEM/F12 containing 10% FCS to inactivate Trypsin. The cell preparation was then treated with collagenase dissolved in DMEM/F12 (1mg/ml) and incubated at 37 °C for 10 min. The specimens were again triturated and fully dissociated cells were rinsed twice in 500μl of PBS/ 5% FCS and subjected to Fluorescence active cell sorting (FACS).

Fluorescence active cell sorting

We isolated EGFP-positive single cells originating from single cell suspension of the UB via FACS using a Beckman Coulter MoFlo Cell Sorter (tip size 70μm) at the Siteman Cancer Center. The EGFP-positive cells were sorted into 96 well microtiter plates with each well containing 5 μl CellsDirect 2x reaction mix (Invitrogen, PN11753-100). Each microtiter plate contained 56 single cells, four dilution series of 2, 4, 8 and 16 cells, and 2 dilution series of 2, 4, 8 and 100 cells. On each plate 8 wells were kept empty as negative control.

Microfluidic qPCR

Prior to qPCR we specifically amplified the transcripts of 24 genes (Fgf9, Vegfr2, Six2, Wnt4, Gdnf, Sall1, Foxc2, Foxd1, Etv4, Wnt11, Lhx1, Spry1, Ret, Bmp7, Etv5, Hnf1b, Gapdh, Egfp, Pax2, Gfra1, Gata3, Fgfr2 and Wnt9b) according to a protocol from Fluidigm (South San Francisco, CA) (Supplementary Methods). All gene expression levels were then measured by qPCR in duplicate using the Fluidigm BioMark HD platform with a 48.48 Dynamic Arry IFC according to manufacturer's protocol (Supplementary Methods, Fluidigm, PN 100-4109 C1).

Quality control

Using the Fluidigm Real-Time PCR Analysis version 3.1.3 software we obtained Ct values and melting curves for each chamber of each Fluidigm chip in which a qPCR reaction was performed. For each primer set we calculated the median melting curve temperature and we excluded all measurements for subsequent analysis that deviated at least 0.7 °C from that value. In addition we excluded all measurements for further analysis that had Ct values of over 35 (indicative of a failed qPCR). We excluded completely the data for two qPCR

primer sets, those for ETV4 and SHH, which both failed to work. In the case of ETV4, our primers were designed against a splice isoform that is not present in kidney. We also excluded all measurements from one single cell and one two-cell inputs, as these wells were not populated with cells during FACS.

Statistics

All statistical analyses were performed in excel or the R statistical package. ANOVAs and linear models were fit using the “anova” and “lm” functions, interpolation of single cell expression data was performed using “predict”, and hierarchical clustering analyses were performed using the “heatmap” and functions, all in R. P-values were generated for observed features in the hierarchical clustering analysis by bootstrap analysis with 1000 samples, using the “pvclust” package, also in R.

Analysis of promoter kinetics

We converted all individual single cell Ct measurements into real space, assuming the maximal Ct measure possible is 40; in other words we create a new value $Ct' = 2^{(40-Ct)}$. This is important as we want to estimate arithmetic rather than geometric summary statistics. Using these transformed Cts we then calculate means and variances for each gene, and these summary statistics appear as points in panel 2B.

Validation of qPCR measurements

We downloaded seventeen normalized probe intensities from the GUDMAP database from the Affymetrix Mouse 430 2.0 array measurements on E11.5 Ureteric bud “Stalk” and E15.5 Ureteric bud “tip” (Potter lab submissions, GUDMAP) to compare these intensities to the average single gene Ct values of our qPCR measurements.

Immunofluorescence microscopy

Immunofluorescence confocal microscopy was performed essentially as described [9]. Antibodies used were as follows: Ret (primary, goat 1:50 Neuromics GT15002; secondary, Cy5 or Cy3, bovine 1:200), Bmp7 (primary rabbit 1:100, Sigma AV32329-100UG; secondary, Cy3, donkey 1:200), EGFP is described in [9] and Wnt11 (primary, rabbit, 1:100, GeneTex GTX105971; secondary, biotin conjugated donkey antirabbit IgG 1:200 and DyLight 649-conjugated Streptoavidin). All secondary antibodies were from Jackson ImmunoResearch labs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Amanda Knoten for animal care, and Chris Sawyer with technical assistance with the Fluidigm assay. We are grateful to Dr. Rob Mitra for insights and discussions on single cell analysis, and Dr. Reed Cartwright for discussions on phylogenetic analysis methods. Research reported in this publication was partly supported by the NIH grant DK082531 (S. J.). Core services offered by the Genome Technology Access Center (GTAC) are partly supported by the Washington University Institute of Clinical and Translational Sciences grant UL1 TR000448 from the National Center for Advancing Translational Sciences (NCATS) of the NIH. We also thank the Alvin J.

Siteman Cancer Center at Washington University School of Medicine and Barnes-Jewish Hospital in St. Louis, MO., for the use of the Siteman Flow Cytometry Core, which provided the cell sorting service. The Siteman Cancer Center is supported in part by an NCI Cancer Center Support Grant P30 CA91842. The content is solely the responsibility of the authors and does not necessarily represent the official view of the NIH.

Sources of support: Research reported was partly supported by the NIH grant DK082531 (S. J.) and Core services by the Washington University Institute of Clinical and Translational Sciences grant UL1 TR000448 and the Alvin J. Siteman Cancer Center Flow Cytometry Core (P30 CA91842).

References

1. Costantini F, Kopan R. Patterning a complex organ: branching morphogenesis and nephron segmentation in kidney development. Dev Cell. 2010; 18(5):698–712. [PubMed: 20493806]
2. White MA, Myers CA, Corbo JC, Cohen BA. Massively parallel in vivo enhancer assay reveals that highly local features determine the cisregulatory function of ChIP-seq peaks. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110(29):11952–11957. [PubMed: 23818646]
3. Gerstein MB, Kundaje A, Hariharan M, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012; 489(7414):91–100. [PubMed: 22955619]
4. Spurgeon SL, Jones RC, Ramakrishnan R. High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. PLoS One. 2008; 3(2):e1662. [PubMed: 18301740]
5. Schuchardt A, D'Agati V, Larsson-Blomberg L, et al. Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret. Nature. 1994; 367(6461):380–383. [PubMed: 8114940]
6. Jain S, Naughton CK, Yang Met al. Mice expressing a dominant-negative Ret mutation phenocopy human Hirschsprung disease and delineate a direct role of Ret in spermatogenesis. Development. 2004; 131(21):5503–5513. [PubMed: 15469971]
7. Jain S, Encinas M, Johnson EM Jr, Milbrandt J. Critical and distinct roles for key RET tyrosine docking sites in renal development. Genes Dev. 2006; 20(3):321–333. [PubMed: 16452504]
8. Jijiwa M, Fukuda T, Kawai K, et al. A targeting mutation of tyrosine 1062 in Ret causes a marked decrease of enteric neurons and renal hypoplasia. Mol Cell Biol. 2004; 24(18):8026–8036. [PubMed: 15340065]
9. Hoshi M, Batourina E, Mendelsohn C, Jain S. Novel mechanisms of early upper and lower urinary tract patterning regulated by RetY1015 docking tyrosine in mice. Development. 2012; 139(13): 2405–2415. [PubMed: 22627285]
10. Harding SD, Armit C, Armstrong J, et al. The GUDMAP database--an online resource for genitourinary research. Development. 2011; 138(13):2845–2853. [PubMed: 21652655]
11. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 2008; 135(2):216–226. [PubMed: 18957198]
12. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005; 123(6):1025–1036. [PubMed: 16360033]
13. Raj A, Peskin CS, Tranchina D, et al. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 2006; 4(10):e309. [PubMed: 17048983]
14. Corish P, Tyler-Smith C. Attenuation of green fluorescent protein half-life in mammalian cells. Protein Eng. 1999; 12(12):1035–1040. [PubMed: 10611396]
15. Barrow J, Bernardo AS, Hay CW, et al. Purification and Characterization of a Population of EGFP-Expressing Cells from the Developing Pancreas of a Neurogenin3/EGFP Transgenic Mouse. Organogenesis. 2005; 2(1):22–27. [PubMed: 19521525]

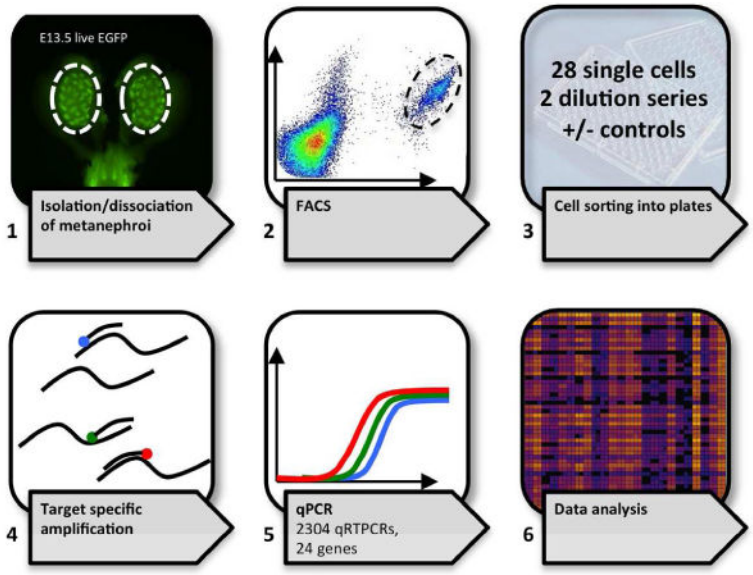

Figure 1.
Overview of the study.

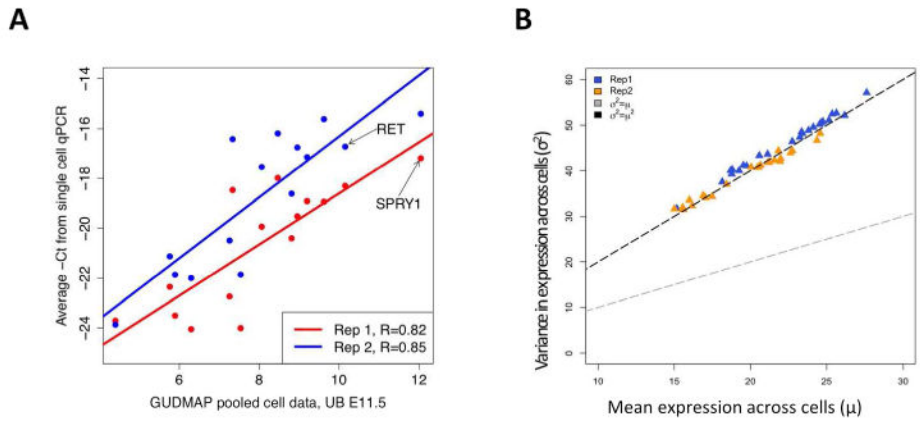

Figure 2.

Expression dynamics at the single cell level. (A) When we compared the average single-cell expression levels measured in our experiments to those obtained by pooled cell analysis of E11.5 UB in GUDMAP, we observed remarkable concordance between the two: the correlation coefficients were 0.82 and 0.85, for run 1 and run 2 respectively. The average single-cell expression levels showed lower correlation to E15.5 ureteric tip (data not plotted, R=0.6 and R=0.55). (B) The nature of gene expression variation at the single-cell level is an active area of experimental and theoretical research [11]. A simple constant-rate model of transcription predicts a linear scaling between the mean and variance of expression levels (the light grey hashed line, $\sigma^2 = \mu$), while a model of punctuated, explosive, "burst-like", transcription predicts a quadratic scaling (the black hashed line, $\sigma^2 = \mu^2$). We observed quadratic scaling between the mean and variance of all genes in our single-cell expression measurements, consistent with "bursty" promoter kinetics. Here, we use a single point to represent the observed mean and variance of transformed CT values for a single gene, calculated across all 28 single cells (Methods). Blue triangles, first technical replicate; yellow triangles, second technical replicate).

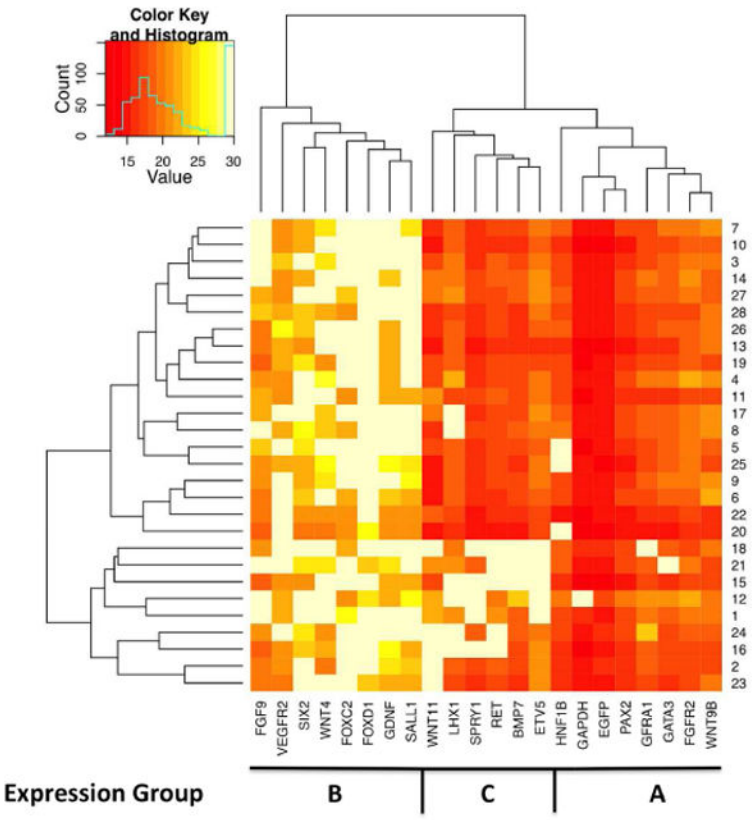

Figure 3.

Hierarchical clustering of single cell qPCR measurements shows population structure. We performed hierarchical clustering on 22 single gene qPCR measurements generated from 28 single cells, averaging over replicates of primer sets and plates, producing a matrix of 616 values (Methods). The data show three gene clusters and two cell clusters. The three gene clusters correspond to genes with consistently (A) high expression, (B) consistently low expression, or (C) variable expression across cells. A surprising result was discordant expression of genes that are typically thought to act in concert. For example, Wnt11 and Etv5 are considered downstream targets of Ret and expected to have concordant expression. While this is true in most of the single cells in this group, however, we saw in few cells where their expression was inversely related (Figure 3, cells 1, 2, 12, 15, 16, 21, 23 and 24). What these variations mean or their effect on the net phenotype of a particular sequence of events from an immature UB tip cell to a more differentiated or daughter cell requires further studies.

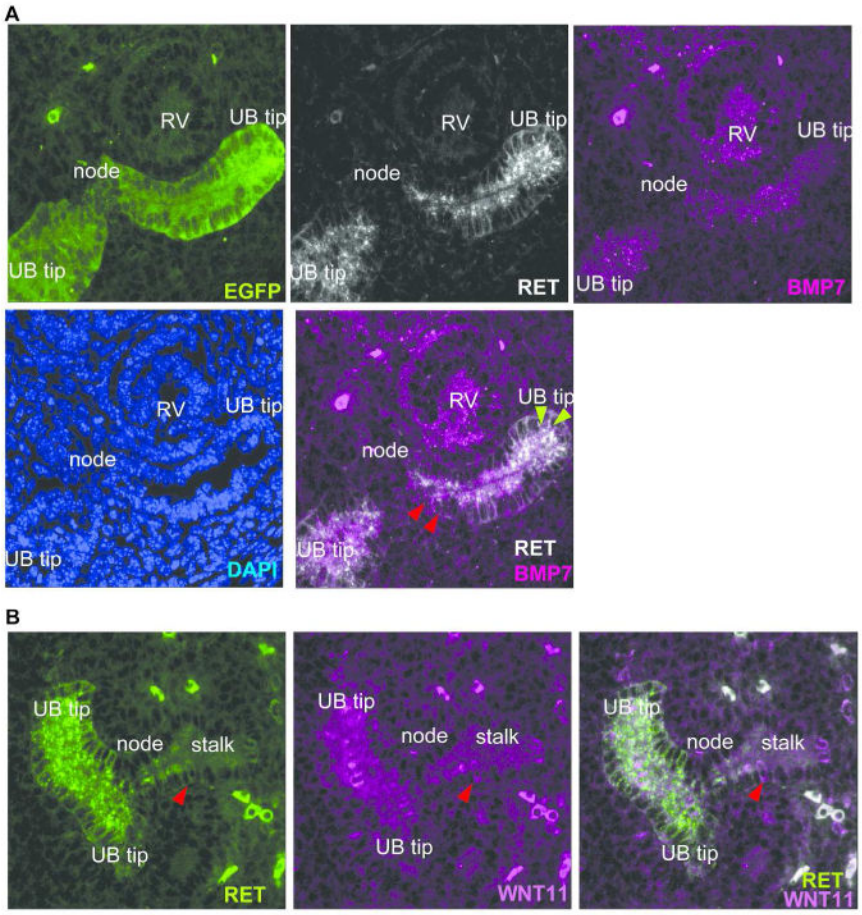

Figure 4.

Expression of Ret, Bmp7 and Wnt11 in the ureteric bud. Images from confocal immunofluorescence microscopy to localize the indicated markers are shown from E13.5 Ret-EGFP mice. (A) Ret and Bmp7 are coexpressed in majority of the cells in the UB tip, but a clear gradient of Ret expression can be seen in the UB tip (green arrows, high near tip; red arrows, low near the node) Bmp7 is expressed throughout the UB and also in the renal vesicle-S-shaped body region (RV), consistent with GUDMAP microarray expression data. (B) Ret and Wnt11 are coexpressed at high levels in the UB tip with rare cells with relatively low Ret and high Wnt11 expression (red arrows).

Table 1

Variance components of the single cell qPCR measurements

	df	Sum sq	Mean Sq	F value	Pr (>F)
Cell	21	1317.7	62.75	8.23	<2 e-16
Gene	1	1190.2	1190.24	156	<2 e-16
Runs	1	14.8	14.83	1.95	0.163
Residuals	1924	14676.7	7.63		